22

УЧЕБНИКИ МГУ
им. М.В.Ломоносова

О.О. Замков, Ю.А. Черемных
А.В. Толстопятенко

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ

2-издание

ДИС

Учебники Московского государственного университета
имени М.В. Ломоносова

О.О. Замков, А.В. Толстопятенко, Ю.Н. Черемных

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ

Учебник

*Под общей редакцией профессора,
доктора экономических наук Сидоровича А.В.*

2-е издание

Москва
Издательство «Дело и Сервис»
1999

ББК 22.1
З-26

Московский государственный университет имени М.В. Ломоносова

Кафедра экономической теории Института переподготовки и повышения квлификации преподавателей социальных и гуманитарных наук

Совместный Центр переподготовки Московского государственного университета и Института экономического развития Всемирного банка

Серия «Учебники МГУ им. М.В. Ломоносова»

Под общей редакцией профессора, д. э. н. А.В. Сидоровича

Замков О.О., Толстопятенко А.В., Черемных Ю.Н.
З-26 Математические методы в экономике: Учебник. 2-е изд. – М.: МГУ им. М.В. Ломоносова, Издательство «Дело и Сервис», 1999. – 368 с.

ISBN 5-86509-054-2

В учебнике дано систематическое изложение основных базовых математических методов, используемых в экономике. Он содержит не только инструментарий математического анализа, знание которого необходимо любому грамотному экономисту, но и многочисленные примеры его применения. Отличительная особенность учебника состоит в том, что он соединяет изучение математических методов с содержательным рассмотрением разделов экономики.

Учебник предназначен для тех, кто изучает экономику, а также сталкивается со статистическими и экономическими моделями в экономике.

ISBN 5-86509-054-2 ББК 22.1

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Вашему вниманию предлагается учебник, который выходит в серии “Учебники Московского университета”. Данная серия, включающая “Курс экономической теории”, учебное пособие по микроэкономике, а также учебники по макроэкономике и математическим методам в экономике дает целостное освещение всех основных разделов современной экономической теории.

Современная экономическая наука характеризуется широким использованием математики. Математические методы стали составной частью методов любой экономической науки, включая экономическую теорию. Ее использование в единстве с обстоятельным экономическим анализом открывает новые возможности для экономической науки и практики. Однако, к сожалению, учебники и учебные пособия, в которых анализируются математические методы в связи с содержательным анализом экономики, практически отсутствуют. Данный учебник восполняет пробел в этой области.

Особенности учебника. Учебник содержит не только инструментарий экономического анализа, знание которого необходимо любому грамотному экономисту, но и многочисленные примеры его применения. Последние способствуют лучшему использованию экономических знаний при выборе математического инструментария и построении экономико-математических моделей. Вопросы, приведенные в конце каждой главы, помогают читателю лучше разобраться в пройденном материале и проверить его усвоение.

Другая особенность состоит в том, что в книге читатель найдет как общую методологию использования математического инструментария и математических моделей в экономике, так и конкретное изложение основных математических понятий и методов: функций и графиков функциональных зависимостей, производных и эластичности, предельного анализа и направлений его применения в экономике (построение и анализ функциональных зависимостей и решение оптимизационных задач различной сложности), понятий и методов теории игр, понятий и методов теории вероятностей, математической статистики и эконометрики с многочисленными примерами их применения.

Рекомендации по использованию. Данный учебник особенно полезен тем, кто изучает микроэкономику на продвинутом уровне, по учебникам, насыщенным математическими моделями, а также сталкивается со статистическими и эконометрическими моделями в экономике. Он приучает читателя к работе с моделями, продвигая его от понимания того, как устроены простейшие модели в экономике, к пониманию более сложных и совершенных экономико-математических моделей.

Этот учебник может удовлетворить потребности как студентов, так и преподавателей экономики в дефицитных сейчас учебниках такого рода. Его отличает широкий охват материала, изложенного на достаточно простом, часто прикладном уровне, призванном восполнить пробелы экономистов в математическом образовании. Предлагаемый учебник может быть также использован для повторения давно пройденного и поэтому забытого материала и для первоначального знакомства тех, кто только начинает изучать экономику, с математическими методами, используемыми в экономике.

Авторы. Авторы учебника - профессор Ю.Н. Черемных, доценты О.О. Замков и А.В. Толстопятенко. Они работают в Московском университете на кафедре экономической теории ИППК МГУ и на экономическом факультете МГУ, а также преподают в Совместном центре переподготовки Московского университета и Института экономического развития Всемирного Банка. Глава 13 написана О.Ю. Шибалкиным.

Выражаем благодарность А.О. Вереникину за подготовку данного пособия к печати.

Мы будем благодарны читателям за отзывы об учебнике, которые просим направлять по адресу: 119899, Москва, Воробьевы горы, МГУ, 2-й корпус гуманитарных факультетов, ИППК, кафедра экономической теории.

А.В. Сидорович, доктор экономических наук, профессор, заведующий кафедрой экономической теории ИППК МГУ, редактор серии учебников Московского университета по экономике.

ВВЕДЕНИЕ

Эта книга написана сотрудниками МГУ им.М.В.Ломоносова и Института экономического развития Всемирного Банка. Использованные в ней материалы подготовлены в ходе реализации пятилетней программы переподготовки преподавателей и государственных служащих России и других стран СНГ в области теории рыночной экономики, организованной ИЭР Всемирного Банка.

Необходимость курса "Математические методы для экономистов" в данной программе была обусловлена тем, что современная экономическая теория предполагает существенно более высокий уровень формализации, чем это было принято в отечественной высшей школе. В то же время литература на русском языке, которая могла бы быть эффективно использована для программ данного типа, практически отсутствует.

Главы 1-12 написаны О.О.Замковым, А.В.Толстопятенко и Ю.Н.Черемных по материалам лекций, прочитанных ими в программах ИЭР Всемирного Банка. Глава 13 "Введение в математическую теорию игр" написана О.Ю.Шибалкиным. Главы 14-19 написаны О.О.Замковым и А.В.Толстопятенко.

В книге содержатся основы знаний по математике, необходимые для экономистов. Это методы построения графиков, исследования функций одной и нескольких переменных, нахождения безусловных и условных экстремумов. Все изложение материала в соответствующих главах ориентировано на задачи экономической теории, прежде всего - микроэкономики.

В других главах непосредственно излагаются отдельные разделы микроэкономики, но акцент при этом делается на математическом обосновании соответствующих утверждений. При этом предполагается, что связанные с этим вопросы содержательного характера обсуждаются в основном в курсе микроэкономики, полностью скоординированном и читаемом параллельно с курсом математических методов. Здесь имеются в виду главы и разделы по теории потребительского выбора, производственным функциям, задачам оптимизации производства, моделированию экономической динамики, статистическому оцениванию макроэкономических зависимостей.

Наконец, некоторые главы посвящены частным математическим методам, особенно широко и непосредственно применяемым в различных разделах экономической теории. Это методы анализа эластичности, связи суммарных, средних и предельных величин, игровые методы, оценивание и анализ уравнений линейной регрессии. В этих главах рассказывается об основных приложениях соответствующих методов в экономике, даются различные примеры.

Авторы выражают благодарность А.О.Вереникину за подготовку данной работы к печати.

ГЛАВА 1
ВВЕДЕНИЕ В МАТЕМАТИЧЕСКИЕ МЕТОДЫ

1.1. Моделирование в экономике и его использование в развитии и формализации экономической теории

Современная экономическая теория, как на микро-, так и на макроуровне, включает как естественный, необходимый элемент математические модели и методы. Использование математики в экономике позволяет, во-первых, выделить и формально описать наиболее важные, существенные связи экономических переменных и объектов: изучение столь сложного объекта предполагает высокую степень абстракции. Во-вторых, из четко сформулированных исходных данных и соотношений методами дедукции можно получать выводы, адекватные изучаемому объекту в той же мере, что и сделанные предпосылки. В-третьих, методы математики и статистики позволяют индуктивным путем получать новые знания об объекте: оценивать форму и параметры зависимостей его переменных, в наибольшей степени соответствующие имеющимся наблюдениям. Наконец, в-четвертых, использование языка математики позволяет точно и компактно излагать положения экономической теории, формулировать ее понятия и выводы.

Математические модели использовались с иллюстративными и исследовательскими целями еще Ф.Кенэ (1758 г., "Экономическая таблица"), А.Смитом (классическая макроэкономическая модель), Д.Рикардо (модель международной торговли). В XIX веке большой вклад в моделирование рыночной экономики внесла математическая школа (Л.Вальрас, О.Курно, В.Парето, Ф.Эджворт и др.). В XX веке математические методы моделирования применялись очень широко, с их использованием связаны практически все работы, удостоенные Нобелевской премии по экономике (Д.Хикс, Р.Солоу, В.Леонтьев, П.Самуэльсон и др.). Развитие микроэкономики, макроэкономики, прикладных дисциплин связано со все более высоким уровнем их формализации. Основу для этого заложил прогресс в области прикладной математики - теории игр, математического программирования, математической статистики. В России в начале XX века большой вклад в математическое моделирование экономики внесли В.К.Дмитриев и Е.Е.Слуцкий. В 1930-е - 50-е годы в этой области не наблюдалось прогресса вследствие идеологических ограничений тоталитарного режима. В 1960-е - 80-е годы экономико-математическое направление возродилось (В.С.Немчинов, В.В.Новожилов, Л.В.Канторович), но было связано в основном с попытками формально описать "систему оптимального функционирования социалистической экономики" (СОФЭ, Н.П.Федоренко, С.С.Шаталин и др.). Строились многоуровневые системы моделей народнохозяйственного планирования, оптимизационные модели отраслей и предприятий. Сейчас важной задачей является моделирование процессов переходного периода.

Любое экономическое исследование всегда предполагает объединение теории (экономической модели) и практики (статистических данных). Мы используем теоретические модели для описания и объяснения наблюдаемых процессов и собираем статистические данные с целью эмпирического построения и обоснования моделей.

Экономические модели. Понятие экономической модели.

Для изучения различных экономических явлений экономисты используют их упрощенные формальные описания, называемые экономическими моделями. Примерами экономических моделей являются модели потребительского выбора, модели фирмы, модели экономического роста, модели равновесия на товарных, факторных и финансовых рынках и многие другие. Строя модели, экономисты выявляют существенные факторы, определяющие исследуемое явление и отбрасывают детали, несущественные для решения поставленной проблемы. Формализация основных особенностей функционирования экономических объектов позволяет оценить возможные последствия воздействия на них и использовать такие оценки в управлении.

Как обычно строится экономическая модель?

1. Формулируются предмет и цели исследования.
2. В рассматриваемой экономической системе выделяются структурные или функциональные элементы, соответствующие данной цели, выявляются наиболее важные качественные характеристики этих элементов.
3. Словесно, качественно описываются взаимосвязи между элементами модели.
4. Вводятся символические обозначения для учитываемых характеристик экономического объекта и формализуются, насколько возможно, взаимосвязи между ними. Тем самым, формулируется математическая модель.
5. Проводятся расчеты по математической модели и анализ полученного решения.

Математическая структура модели и ее содержательная интерпретация

Следует различать математическую структуру модели и ее содержательную интерпретацию. Рассмотрим следующие два простых примера.

Пример 1. Пусть требуется определить, какую сумму следует положить в банк при заданной ставке процента (20% годовых), чтобы через год получить \$12000?

Вводя формальные обозначения для величин, фигурирующих в задаче:

начальная сумма денег - M_0,
конечная сумма денег - M_1,
ставка процента - R
и записывая соотношение между ними

$$M_1 = M_0 \cdot \left(1 + \frac{R}{100}\right),$$

найдем требуемую величину из решения основного уравнения модели

$$M_0 = \frac{M_1}{1 + \frac{R}{100}} = \frac{\$12000}{1{,}2} = \$10000.$$

Пример 2. Пусть требуется определить, каков был объем выпуска продукции завода, если в результате технического перевооружения средняя производительность труда увеличилась на 20%, и завод стал выпускать 12000 единиц продукции.

Вводя формальные обозначения для величин, фигурирующих в задаче:

начальный выпуск - Q_0,
конечный выпуск - Q_1,
процент прироста производительности - R,
и записывая соотношение между ними (следующее из определения средней производительности труда $\frac{Q}{L}$)

$$Q_1 = Q_0 \frac{L_1}{L_0} = Q_0 \left(1 + \frac{(L_1 - L_0)}{L_0}\right) = Q_0 (1 + \frac{R}{100}),$$

найдем искомую величину из решения основного уравнения модели

$$Q_0 = \frac{Q_1}{1 + \frac{R}{100}} = \frac{\$12000}{1{,}2} = \$10000.$$

Сравнивая полученные модели и результаты, мы можем заметить, что математическая форма модели

$$X_1 = X_0 \cdot \left(1 + \frac{R}{100}\right)$$

и даже числовые значения входящих в нее величин в обоих случаях одинаковы, однако экономическая ситуация, описываемая моделью, экономическая интерпретация модели и результатов расчета совершенно различны. Таким образом, одни и те же математические модели и методы могут быть использованы для решения совершенно различных экономических задач.

Роль моделей в экономической теории и принятии решений

Экономические модели позволяют выявить особенности функционирования экономического объекта и на основе этого предсказывать будущее поведение объекта при изменении каких-либо параметров. Предсказание будущих изменений, например, повышение обменного курса, ухудшение экономической конъюнктуры, падение прибыли может опираться лишь на интуицию. Однако при этом могут быть упущены, неправильно определены или неверно оценены важные взаимосвязи экономических показателей, влияющие на рассматриваемую ситуацию. В модели все взаимосвязи переменных могут быть оценены количественно, что позволяет получить более качественный и надежный прогноз.

Для любого экономического субъекта возможность прогнозирования ситуации означает, прежде всего, получение лучших результатов или избежание потерь, в том числе и в государственной политике.

Неполнота экономической модели

По своему определению любая экономическая модель абстрактна и, следовательно, неполна, поскольку выделяя наиболее существенные факторы, определяющие закономерности функционирования рассматриваемого экономического объекта, она абстрагируется от других факторов, которые, несмотря на свою относительную малость, все же в совокупности могут определять не только отклонения в поведении объекта, но и само его поведение. Так, в простейшей модели спроса считается, что величина спроса на какой-либо товар определяется его ценой и доходом потребителя. На самом же деле на величину спроса оказывает также влияние ряд других факторов: вкусы и ожидания потребителей, цены на другие товары, воздействие рекламы, моды и так далее. Обычно предполагают, что все факторы, не учтенные явно в экономической модели, оказывают на объект относительно малое результирующее воздействие в интересующем нас аспекте. Состав учтенных в модели факторов и ее структура могут быть уточнены в ходе совершенствования модели.

1.2. Математическая модель и ее основные элементы. Экзогенные и эндогенные переменные, параметры. Виды зависимостей экономических переменных и их описание. Уравнения, тождества, неравенства и их системы

Математическая модель экономического объекта - это его гомоморфное отображение в виде совокупности уравнений, неравенств, логических отношений, графиков. Гомоморфное отображение объединяет группы отношений элементов изучаемого объекта в аналогичные отношения элементов модели. Иными словами, модель - это условный образ объекта, построенный для упрощения его исследования. Предполагается, что изучение модели дает новые знания об объекте, либо позволяет определить наилучшие решения в той или иной ситуации.

Для описания основных видов элементов экономической модели рассмотрим конкретную ситуацию и построим соответствующую ей модель.

Пусть имеется фирма, выпускающая несколько видов продукции. В процессе производства используются три вида ресурсов: оборудование, рабочая сила и сырье; эти ресурсы однородны, количества их известны и в данном производственном цикле увеличены быть не могут. Задан расход каждого из ресурсов на производство единицы продукции каждого вида. Заданы цены продуктов. Нужно определить объемы производства с целью максимизации стоимости произведенной продукции (или, в предположении, что вся она найдет сбыт на рынке - общей выручки от реализации).

Для решения поставленной задачи нужно построить математическую модель, наполнить ее информацией, а затем провести по ней необходимые расчеты. Вначале при построении модели нужно определить индексы, экзогенные и эндогенные переменные и параметры. В нашей задаче свой индекс должен иметь каждый вид продукции (пусть это индекс i, меняющийся от 1 до n), а также вид ресурсов (если мы обозначаем их одной переменной; пусть в нашей задаче ресурсы обозначены разными переменными). Далее опишем экзогенные переменные - те, которые задаются вне модели, т.е. известны заранее, и параметры - это коэффициенты уравнений модели. Часто экзогенные переменные и параметры в моделях не разделяют. В рассматриваемой задаче заданы экзогенные переменные - имеющиеся количества оборудования K, рабочей силы L и сырья R; заданы параметры - коэффициенты их расхода на единицу i-й продукции k_i, l_i и r_i соответственно. Цены продуктов p_i также известны.

Далее вводятся обозначения для *эндогенных* переменных - тех, которые определяются в ходе расчетов по модели и не задаются в ней извне. В нашем случае - это неизвестные объемы производства продукции каждого i-го вида; обозначим их x_i.

Закончив описание переменных и параметров, переходят к формализации условий задачи, к описанию ее допустимого множества и целевой функции (если таковая имеется). В нашей задаче допустимое множество - это совокупность всех вариантов производства, обеспеченных имеющимися ресурсами. Оно описывается с помощью системы неравенств:

$$\begin{array}{c} k_1x_1+k_2x_2+\ldots+k_nx_n\le K, \\ l_1x_1+l_2x_2+\ldots+l_nx_n\le \bar{L}, \\ r_1x_1+r_2x_2+\ldots r_nx_n\le R, \end{array} \quad \text{или} \quad \begin{array}{c} \sum\limits_i k_ix_1\le K, \\ \sum\limits_i l_ix_i\le L, \\ \sum\limits_i r_ix_i\le R. \end{array}$$

К этим ограничениям по ресурсам добавляются требования неотрицательности переменных $x_i \ge 0$. Если бы какой-то ресурс нужно было израсходовать полностью (например, полностью занять всю рабочую силу), соответствующее неравенство превратилось бы в уравнение. Это сузило бы допустимое множество и, возможно, исключило бы из него первоначально наилучшее решение.

Если модель является оптимизационной (а данная модель такова), то наряду с ограничениями должна быть выписана целевая функция, т.е. максимизируемая или минимизируемая величина, отражающая интересы принимающего решение субъекта. Для данной задачи максимизируется величина

$$p_1x_1 + p_2x_2 + \ldots + p_nx_n, \quad \text{или} \quad \sum_i p_i \cdot x_i \to \max.$$

Поставленная задача далеко не всегда хорошо описывает ситуацию и соответствует задачам лица, принимающего решение (ЛПР). В действительности, по крайней мере:

1) ресурсы до некоторой степени взаимозаменяемы;

2) затраты ресурсов не строго пропорциональны выпуску (есть постоянные затраты, не связанные с объемом выпуска; предельные затраты меняются);

3) объемы ресурсов не строго фиксированы, они могут покупаться и продаваться, браться или сдаваться в аренду;

4) внутри каждого вида ресурсов можно выделить составляющие, функционально или качественно различные, в той или иной мере заменяющие или дополняющие друг друга и по-разному влияющие на объем выпуска;

5) цена продукта может зависеть от объема его реализации, то же касается цены ресурса;

6) фирма может использовать одну из конечного набора технологий (или сочетание нескольких таких технологий), характеризующихся определенными сочетаниями используемых ресурсов;

7) различные единицы получаемой прибыли могут иметь разную ценность для лица, принимающего решение (что обусловлено, например, особенностями налоговой системы);

8) интересы и предпочтения субъекта не ограничиваются максимизацией объема прибыли, поэтому целевая функция должна учитывать и другие количественные и качественные показатели;

9) для субъекта реально решаемая задача не ограничивается одним моментом или периодом времени, важны динамические взаимосвязи;

10) на ситуацию могут воздействовать случайные факторы, которые необходимо принять во внимание.

Многие разделы экономической теории посвящены изучению, описанию и моделированию перечисленных аспектов на различных уровнях хозяйственной деятельности, с той или иной степенью детализации и в различных сочетаниях.

1.3. Основные типы моделей

Математические модели, используемые в экономике, можно подразделять на классы по ряду признаков, относящихся к особенностям моделируемого объекта, цели моделирования и используемого инструментария: модели макро- и микроэкономические, теоретические и прикладные, оптимизационные и равновесные, статические и динамические.

Макроэкономические модели описывают экономику как единое целое, связывая между собой укрупненные материальные и финансовые показатели: ВНП, потребление, инвестиции, занятость, процентную ставку, количество денег и другие. ***Микроэкономические*** модели описывают взаимодействие структурных и функциональных составляющих экономики, либо поведение отдельной такой составляющей в рыночной среде. Вследствие разнообразия типов экономических элементов и форм их взаимодействия на рынке, микроэкономическое моделирование занимает основную часть экономико-математической теории. Наиболее серьезные теоретические результаты в микроэкономическом моделировании в последние годы получены в исследовании стратегического поведения фирм в условиях олигополии с использованием аппарата теории игр.

Теоретические модели позволяют изучать общие свойства экономики и ее характерных элементов дедукцией выводов из фор-

1405

мальных предпосылок. ***Прикладные*** модели дают возможность оценить параметры функционирования конкретного экономического объекта и сформулировать рекомендации для принятия практических решений. К прикладным относятся прежде всего эконометрические модели, оперирующие числовыми значениями экономических переменных и позволяющие статистически значимо оценивать их на основе имеющихся наблюдений.

В моделировании рыночной экономики особое место занимают ***равновесные*** модели. Они описывают такие состояния экономики, когда результирующая всех сил, стремящихся вывести ее из данного состояния, равна нулю. В нерыночной экономике неравновесие по одним параметрам (например, дефицит) компенсируется другими факторами (черный рынок, очереди и т.п.). Равновесные модели дескриптивны, описательны. В нашей стране долгое время преобладал нормативный подход в моделировании, основанный на ***оптимизации***. Оптимизация в теории рыночной экономики присутствует в основном на микроуровне (максимизация полезности потребителем или прибыли фирмой); на макроуровне результатом рационального выбора поведения экономическими субъектами оказывается некоторое состояние равновесия.

В моделях ***статических*** описывается состояние экономического объекта в конкретный момент или период времени; ***динамические*** модели включают взаимосвязи переменных во времени. В статических моделях обычно зафиксированы значения ряда величин, являющихся переменными в динамике, - например, капитальных ресурсов, цен и т.п. Динамическая модель не сводится к простой сумме ряда статических, а описывает силы и взаимодействия в экономике, определяющие ход процессов в ней. Динамические модели обычно используют аппарат дифференциальных и разностных уравнений, вариационного исчисления.

Детерминированные модели предполагают жесткие функциональные связи между переменными модели. ***Стохастические*** модели допускают наличие случайных воздействий на исследуемые показатели и используют инструментарий теории вероятностей и математической статистики для их описания.

1.4. Математическая экономика и эконометрика

Математическая экономика - раздел экономической науки, занимающийся анализом свойств и решений математических моделей экономических процессов. В некоторых случаях эти модели могут рассматриваться как часть математической теории на стыке с экономической наукой. Математическая экономика отделяется обычно от эконометрики, занимающейся статистической оценкой и анализом экономических зависимостей и моделей на основе изучения эмпирических данных. В математической экономике исследуются теоретические модели, основанные на определенных формальных предпосылках (линейность, выпуклость, монотонность и т.п. зависимости, конкретные формулы взаимосвязи величин). Математическая экономика, вообще говоря, не занимается изучением степени обоснованности того, что данная зависимость имеет тот или иной вид (например, что величина потребления является линейной возрастающей функцией дохода), - это оставляется для эконометрики. Задачей математической экономики является изучение вопроса о существовании решения модели, условиях его неотрицательности, стационарности, наличия других свойств. Это обычно осуществляется, как и в математике, путем дедуктивного получения следствий (теорем) из априорно сделанных предпосылок (аксиом).

Разумеется, предметная область, методология и инструментарий экономической науки не исчерпываются подходами математической экономики и эконометрики - обычно в экономических исследованиях используются также методы качественного анализа, индуктивные, эвристические подходы, перемежающиеся с элементами математической экономики и эконометрики. Таким образом, математическая экономика выступает и как самостоятельный раздел экономической науки, и как один из ее инструментов. При этом разделы математической экономики, исследовавшиеся ранее в чисто теоретическом плане, все больше становятся теоретической базой и элементами прикладных исследований.

Среди моделей математической экономики можно выделить два крупных класса - модели равновесия в экономических системах и модели экономического роста. Модели равновесия (например, модель Эрроу-Дебре, модель "затраты - выпуск" В.Леонтьева) помогают исследовать состояния экономических систем, в которых равнодействующая всех внешних сил равна нулю. Это, вообще говоря, статические модели, в то время как экономическая динамика описывается с помощью моделей роста (модель Харрода-Домара, модель Солоу, модели магистрального типа и др.). Ключевым моментом исследования моделей роста является анализ и отыскание траекторий стационарного роста (роста с постоянными, в том или ином смысле, структурными характеристиками), к выходу на которые обычно стремится описываемая моделью экономическая система. Исследование траекторий стационарного роста является одновременно базой для анализа более сложных типов роста и связующим звеном с моделями экономического равновесия (поскольку отыскание такой траектории равнозначно отысканию меняющегося вполне определенным образом равновесного состояния). Значительный вклад в теорию роста внесли работы фон Неймана, Солоу, Гейла, Моришимы и др.

Эконометрика - наука, исследующая количественные закономерности и взаимозависимости в экономике при помощи методов математической статистики. Основа этих методов - корреляционно-регрессионный анализ. Использование современных методов математической статистики началось в биологии. В последней четверти XIX века английский биолог К.Пирсон положил начало современной математической статистике изучением кривых распределения числовых характеристик человеческого организма. Затем он и его школа перешли к изучению корреляций в биологии и построению линейных функций регрессии.

Первые работы по эконометрике появились в конце XIX - начале XX века. В 1897 г. появилась работа одного из основателей математической школы в экономической теории В.Парето, посвященная статистическому изучению доходов населения в разных странах. Была предложена кривая Парето $y = A(x-a)^{-\alpha}$, где x -величина дохода; y - численность лиц, имеющих доход, больший x; a - минимальный доход; A и α - параметры зависимости, получаемые статистическими методами.

В самом начале XX века вышло несколько работ английского статистика Гукера, в которых он применил корреляционно-регрессионные методы, разработанные Пирсоном и его школой, для изучения взаимосвязей экономических показателей, в частности - влияния числа банкротств на товарной бирже на цену зерна. В работах Гукера содержалась идея временного лага между экономическими переменными, а также идея корреляционного анализа не самих величин, а их приращений. В дальнейшем появилось огромное число работ как по развитию теории математической статистики и ее прикладных элементов, так и по практическому приложению этих методов в экономическом анализе. К первой группе могут быть, например, отнесены работы Р.Фишера по дисперсионному анализу, ко второй - работы по оценке и исследованию производственных функций, в частности - классическая работа Кобба и Дугласа 1928 г.

Эконометрические модели и методы сейчас - это не только мощный инструментарий для получения новых знаний в экономике, но и широко применяемый аппарат для принятия практических решений в прогнозировании, банковском деле, бизнесе.

Вопросы к главе 1

1. Почему необходимо использование математики в экономике?
2. Что такое математическая модель?
3. Как строится математическая модель экономического явления или объекта? Приведите пример построения и уточнения модели.
4. Какова связь между математической структурой модели и ее содержательной интерпретацией?
5. Какие переменные модели называются экзогенными, а какие - эндогенными?
6. Пусть в модели, приведенной в лекции, всё сырьё должно быть израсходовано полностью. Как при этом изменится ограничение на сырьё?
7. Если в модели, описанной в главе, предположить, что количество рабочей силы может меняться, то придется изменить ограничения или целевую функцию?
8. В модели, описанной в главе, являются ли ресурсы взаимозаменяемыми?
9. Чем отличаются равновесные модели от оптимизационных?
10. В чём отличие статических моделей от динамических? К какому типу относится приведенная в главе модель?
11. В чём отличие математической экономики от эконометрики?
12. Предположим перед вами стоит задача обоснования линейной зависимости объёма потребления от дохода на основе эмпирических данных. Относится ли решение подобной задачи к математической экономике или к эконометрике?
13. Пусть фирма выпускает один вид продукции, причём объём выпуска зависит от затрат только двух факторов: труда и капитала. Заданы цена единицы труда (ставка заработной платы), цена единицы капитала (ставка процента) и издержки фирмы. Задача состоит в определении объёмов затрат труда и капитала, максимизирующих объём выпуска. Постройте соответствующую экономико-математическую модель.
14. В модели, сформулированной в вопросе 13, укажите, какие переменные являются экзогенными, а какие - эндогенными. К какому типу относится эта модель: макроэкономических или микроэкономических?

ГЛАВА 2
ФУНКЦИИ И ГРАФИКИ В ЭКОНОМИЧЕСКОМ МОДЕЛИРОВАНИИ

2.1. Понятие функциональной зависимости. Способы задания и исследования функций

Многочисленные наблюдения и исследования показывают, что в окружающем нас мире величины (например, цена какого-либо товара и величина спроса на этот товар, прибыль фирмы и объем производства этой фирмы, инфляция и безработица и т.п.) существуют не изолированно друг от друга, а напротив, они связаны между собой определенным образом. Понятие функции или функциональной зависимости - одно из основных математических понятий при помощи которых моделируются взаимосвязи между различными величинами, количественные и качественные отношения между различными экономическими характеристиками и показателями.

Понятие функции, как и понятие множества, относится к числу начальных понятий, поэтому оно не определяется, а поясняется. Говорят, что задана *функция* f, если дан закон, согласно которому каждому значению x из некоторого числового множества A ставится в *соответствие* одно вполне определённое значение y из некоторого числового множества B.

Функциональная зависимость между величинами x и y символически обозначается так: $y = f(x)$; говорят, что x - аргумент (независимая переменная), а y - функция (зависимая переменная).

Совокупность всех значений аргумента, каждому из которых соответствует вполне определенное значение функции, называется областью определения функции.

Множество значений, принимаемых y, называется областью изменения функции.

Функцию можно задавать различными способами. Наиболее распространенные и важные среди них - задание функции формулой, таблицей и графиком. При задании функции в ЭВМ часто используется также алгоритмический способ.

В качестве примера рассмотрим взаимосвязь между ценой продукта, которую мы обозначим через р и величиной спроса на этот продукт, которую мы обозначим через q. Эта связь может быть, к примеру, представлена следующей таблицей:

p, руб.	100	150	200	250	300
q, тыс.шт.	18	15	12	9	6

отражающей отрицательную взаимосвязь величин (убывание величины спроса с возрастанием цены).

2.2.1. Построение и анализ графиков функций

Эта же взаимосвязь величин может быть представлена в виде графика на рис. 2.1.

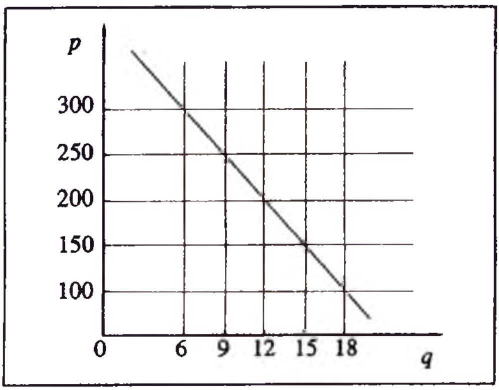


Рис. 2.1

Определение. Графиком функции f называется геометрическое место (множество) точек на координатной плоскости, имеющих координаты $(x;f(x))$, у которых абсциссами служат рассматриваемые значения независимой переменной x, а ординатами - соответствующие значения функции $y=f(x)$

Для того, чтобы построить график функции, имея ее табличное представление, например график функции спроса, достаточно отложить значения величин, приведенных в таблице на соответствующих координатных осях, восстановить перпендикуляры к осям из точек, соответствующих определенному значению цены или спроса, и нанести точки пересечения перпендикуляров.

Функциональная зависимость между величинами x и y может быть задана также в виде формулы $y = f(x)$, в которой в качестве $f(x)$ фигурирует конкретная функция. В данном случае зависимость между ценой и величиной спроса выражается формулой: $p = 500 - 50\frac{q}{3}$ (или $q = 30 - 0{,}06p$). Подставляя в последнюю формулу значения цены, представленные в верхней строке таблицы, мы легко убедимся в том, что в результате получаются соответствующие ценам величины спроса, представленные в нижней строке таблицы. Таким образом, зная формулу функции, несложно получить табличное и графическое представление этой функции.

Свойства функций

Функции характеризуются рядом свойств, к важнейшим из которых (для построения и исследования графиков) относятся: четность, нули, периодичность, монотонность, ограниченность функции, наличие у функций асимптот и обратной функции. Рассмотрим вкратце эти свойства функций.

1. Четные и нечетные функции. Функция $y=f(x)$ называется четной, если для любых двух различных значений аргумента из области ее определения выполняется равенство $f(-x) = f(x)$. Например, $y=x^{2n}$, (n - натуральное); $y=|x|$ - четные функции. Сумма, разность, произведение и частное четных функций есть функция четная.

Функция называется нечетной, если для любого значения аргумента из области определения функции выполняется равенство $f(-x) = -f(x)$. К нечетным функциям относятся, например, $y=x^{2n+1}$, где n - любое натуральное число, $y = \dfrac{x}{x^2 + 4}$ и т.д.

Не всякая функция является либо четной, либо нечетной. Например, функции $y = x^2+3x$; $y = (x + 1)^2$ и т.д. называются аморфными.

Заметим, что если функция $y=f(x)$ четная или нечетная, то область ее определения симметрична относительно центра O. Поэтому функции $y=\sqrt{x}$; $y=\lg x$; $y=a^x$ и т.д. не могут быть ни четными, ни нечетными. Сумма и разность нечетных функций есть функция нечетная, а произведение и частное нечетных функций - функция четная.

График четной функции симметричен относительно оси OY, а нечетной - относительно центра O.

2. Нули функции. Нулями функции $f(x)$ называют те значения аргумента, при которых функция обращается в нуль: $f(x) = 0$. Графически нулями функции являются точки пересечения графика функции с осью абсцисс. Например, нулями функции $y = (x-1)(x+2)$ будут корни уравнения $y = 0$, т.е. $x = 1$ и $x = 2$.

3. Периодические функции. Функция $y=f(x)$ называется периодической, если существует число T такое, что для каждого значения аргумента x из области ее задания имеет место равенство $f(x+T) = f(x)$. Число T называют периодом этой функции. Примеры периодических функций : $y=\sin(x)$; $y=\cos(x)$; (для них наименьший положительный период равен 2р) $y=\text{tg}(x)$; $y=\text{ctg}(x)$ (для них $T =$ р).

4. Монотонные функции. Функция $y=f(x)$ называется возрастающей на некотором промежутке, если для любых значений x из этого промежутка, большему значению аргумента соответствует большее значение функции, т.е. если $x_1 < x_2$, то $f(x_1) < f(x_2)$.

Функция называется убывающей на некотором промежутке, если для любых значений x из этого промежутка, большему значению аргумента соответствует меньшее значение функции, т.е. если $x_1 < x_2$, то $f(x_1) > f(x_2)$.

Как возрастающие, так и убывающие функции называются монотонными функциями.

5. Асимптоты. Асимптотой графика функции называется прямая, к которой сколь угодно близко приближается график данной функции при стремлении аргумента к бесконечности или к некоторому числу a или случай вертикальной асимптоты. Асимптоты могут быть вертикальными, горизонтальными и наклонными.

Вертикальная асимптота - это прямая $x = a$, если $\lim\limits_{x \to a} f(x) = \infty$.

Горизонтальная асимптота - это прямая $y = b$, если $\lim\limits_{x \to \infty} f(x) = b$.

Наклонная асимптота - это прямая $y = kx + b$, если $\lim\limits_{x \to \infty}(f(x) - kx) = b$. Коэффициент наклона k находится путем вычисления предела $\lim\limits_{x \to \infty}\frac{f(x)}{x}$.

6. Ограниченные функции. Функция $f(x)$ называется ограниченной сверху (снизу), если существует такое число M, что для всех x из области определения $f(x) < M$ ($f(x) > M$) (см. рис.2.2в,б).

Функция называется ограниченной, если существует такое число $M > 0$, что для всех x из области определения $|f(x)| < M$ (см. рис.2.2а).

7. Обратная функция и ее график. Дана функция $y=f(x)$, Выразим x как некоторую функцию от y: $x=\varphi(y)$, т.е. представим y как аргумент; x - как функцию. Тогда функция $x=\varphi(y)$ называется обратной по отношению к функции $y = f(x)$, если при подстановке её вместо аргумента f получаем тождественное равенство: $y \equiv f(\varphi(y))$.

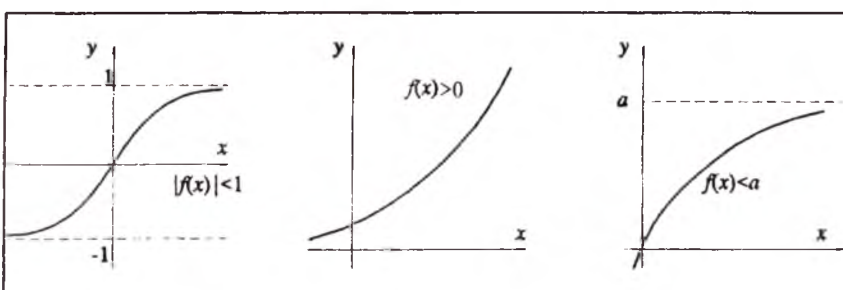


Рис. 2.2а *Рис. 2.2б* *Рис. 2.2в*

Равенство функций называется тождественным, если оно справедливо при всех значениях аргумента из области определения.

Примеры:

1. Дана функция $y = \frac{x}{3} + 1$. Разрешим ее относительно x: $x = 3y - y$. Функция $x = 3y - y$ будет обратной по отношению к функции $y = \frac{x}{3} + 1$, ибо $y = \frac{3y-y}{3} + 1$.

2. Дана функция $y = x^2$ (при $x \leq 0$). Определим x как функцию y: $x = -\sqrt{y}$.

Функция $x = -\sqrt{y}$ - обратная функции $y = x^2$ $(x \leq 0)$.

Отметим, что область определения и область изменения функции $y=f(x)$ и функции, обратной к ней, меняются ролями. Всегда ли существует обратная функция? В этом случае имеет место следующая важная теорема.

Обратная функция существует при $a \quad x \quad b$, если функция $f(x)$ монотонно возрастает или монотонно убывает в интервале $[a,b]$.

8. Сложная функция. Функция, заданная в виде $y = f(g(x))$, называется сложной функцией x или суперпозицией функций g и f. Сложную функцию часто записывают в виде $y = f(u)$, где $u = g(x)$. при этом аргумент x называют независимой переменной, а u - промежуточным аргументом.

9. Неявная функция. Функция, заданная в виде уравнения $F(x,y) = 0$, не разрешенном относительно y, называется неявной функцией x.

Схема исследования функции для построения графика

Перед построением графика функции необходимо провести исследования ее по следующим пунктам:

1. Область определения функции (ООФ),
2. Область изменения функции (ОИФ),
3. Периодичность функции,
4. Четность или нечетность функции,
5. Монотонность функции,
6. Точки пересечения с осями координат:
   а) с осью OY $(x=0)$; б) с осью OX $(y=0)$;
7. Интервалы знакопостоянства,
8. Асимптоты.

2.2.2. Основные элементарные функции и их графики (линейная, степенная, показательная, логарифмическая функции)

1. Линейная функция $y = kx+b$. ООФ $(-\infty,+\infty)$, ОИФ $(-\infty, +\infty)$. График - прямая линия (см. рис.2.3) Угловой коэффициент k равен $\operatorname{tg}\varphi$, где φ - угол между положительным направлением оси Ox и прямой. С увеличением k по абсолютной величине наклон прямой увеличивается. При $k = 0$ имеем: $y = b$ - прямая, параллельная оси абсцисс (Ox). Функция $y = kx+b$ при $k \neq 0$ - монотонная: возрастает при $k > 0$ и убывает при $k < 0$. Возрастающая функция (при $k > 0$) описывает положительную зависимость величин x и y (пример - функция предложения), убывающая функция (при $k < 0$) описывает отрицательную зависимость величин x и y (пример - функция спроса).

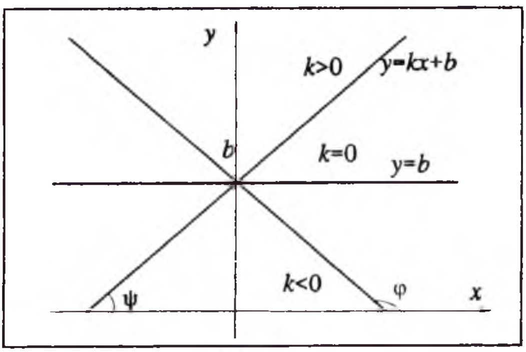


Рис. 2.3

Коэффициент k не только характеризует наклон прямой, но и показывает знак и скорость изменения (возрастания/убывания) функции. Чем больше k, тем быстрее возрастает/убывает функция (т.е. тем больше изменение y при фиксированном изменении x и тем чувствительнее величина y к изменению x). Если $k = 0$, то $y = b = \text{const}$, т.е. величина y постоянна и не зависит от величины x.

Коэффициент b показывает значение функции в точке пересечения графика с осью Oy. Прямая пересекает ось ординат в точке $(0;b)$. Если $b = 0$, то $y = kx$ - это прямая пропорциональная зависимость. Прямая $y = kx$ проходит через начало координат.

2а. Функция $y = \frac{k}{x}$ определена при всех значениях x за исключением точки $x = 0$. ОИФ - интервалы $(-\infty, 0)$, $(0, \infty)$. График - гипербола (см. рис.2.4а).

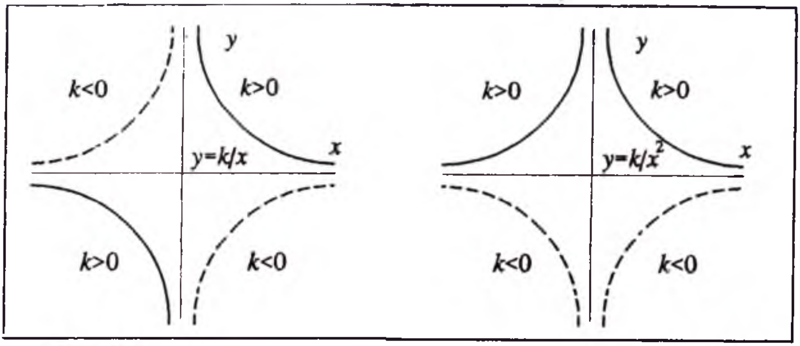


Рис. 2.4а Рис. 2.4б

Функция в каждом из интервалов (-∞, 0) и (0, ∞) монотонная: возрастает при $k < 0$ и убывает при $k > 0$. Она выражает обратную пропорциональную зависимость между x и y. Функция нечетная, следовательно, гипербола симметрична относительно начала координат. Она расположена в первой и третьей четвертях, если $k > 0$ и во второй и четвертой при $k < 0$. Оси координат являются асимптотами гиперболы.

2б. Функция $y = \frac{k}{x^2}$. Эта функция также определена при x - любом, кроме точки x = О, ОИФ - интервалы (-∞, 0),(0, ∞). График - гипербола (см. рис. 2.4б). Функция в интервале (-∞, 0) при $k >$ 0 возрастает, при $k < 0$ убывает; в интервале (0, ∞) при $k > 0$ убывает, при $k < 0$ возрастает. Функция четная. Функция расположена в первой и второй четвертях при $k > 0$ и в третьей и четвертой четвертях при $k < 0$. Оси координат являются асимптотами гиперболы.

3. Показательная функция $y = a^x$ ($a > 0$) определена на всей числовой оси область изменения функции (0,+∞), т.е. ее график находится в верхней полуплоскости. При $a >$ 1 функция монотонно возрастающая, при $a < 1$ функция монотонно убывающая. График проходит через точку (0;1), так как $a^0 = 1$. Ось Ox является асимптотой (см. рис. 2.5). В качестве основания степенной функции часто используется число $e \approx 2{,}71828...$

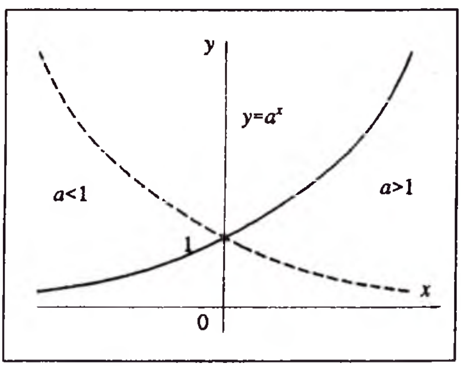


Рис. 2.5

В этом случае функция называется экспонентой.

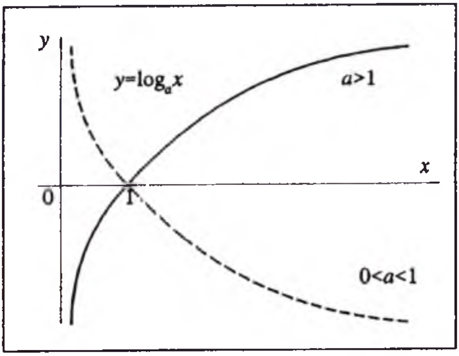


Рис. 2.6

4. Логарифмическая функция $y = \log_a x$ (при $a > 0$ и $a \neq 1$). Функция определена при $x > 0$. ОИФ (-∞, +∞), монотонна (возрастает при $a > 1$, убывает при $a < 1$). График проходит всегда через точку (1;0), так как $\log_a 1 = 0$. Ось ординат является асимптотой для графика (см. рис.2.6). Логарифмическая функция $y = \log_c x$ является обратной по отношению к показательной функции $x = a^y$, так как $\log_a(a^y) \equiv y(a^{(\log_a x)} \equiv x)$. В качестве основания логарифмической функции a часто используется число $e \approx 2{,}71828...$. В этом случае логарифмическая функция называется натуральным логарифмом и обозначается $y = \ln x$.

5. Степенная функция с любым рациональным показателем $y = x^\alpha$. Рассмотренные выше функции: x, x^2, $\frac{1}{x}$, $\frac{1}{x^2}$ являются частными случаями этой функции. В зависимости от б выделяют следующие функции:

5.1. $y = x^{2m}$, где m натуральное число.

Область определения этих функций (-∞, +∞), область значений [0, +∞). Функции четные. Графики этих функций параболы (см. рис.2.7а).

5.2. $y = x^{2m-1}$, где m натуральное число.

Область определения (-∞, +∞), область изменения (-∞, +∞). Функции нечетные. Графики функций также параболы. Во всей области определения функции такого вида - возрастающие (см. рис.2.7б).

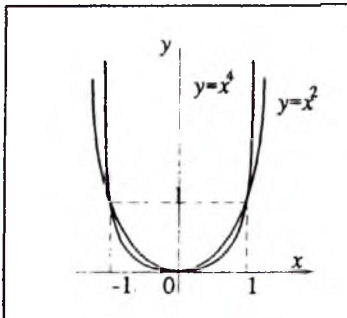


Рис. 2.7а

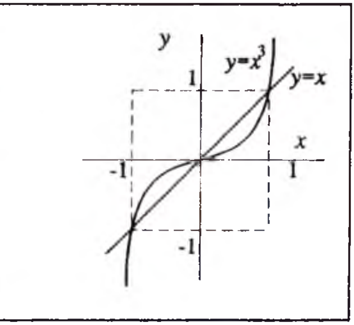


Рис. 2.7б

5.3. $y = x^{-2m}$ $\left[y = \dfrac{1}{x^{2m}}\right]$, где m натуральное число.

Областью определения функций является объединение интервалов $(-\infty, 0)$ и $(0, +\infty)$, а областью значений - множество $(0,+\infty)$. Функции четные. Ветви функций расположены в первой и четвертой четвертях (см. рис.8а) Оси координат являются вертикальной и горизонтальной асимптотами.

5.4. $y = x^{-2m+1}$ $\left[y = \dfrac{1}{x^{2m+1}}\right]$, где m натуральное число.

Областью определения и областью существования функций являются интервалы $(-\infty, 0)$ и $(0, +\infty)$. Функции нечетные. Графики их расположены в первой и третьей четвертях, оси координат также служат асимптотами, на всей области определения функции убывающие. Графики функций называют гиперболами (см. рис.2.8б).

5.5. $y = x^{\alpha}$ $(\alpha > 0)$ (см. рис.2.9а,б)).

Область определения $[0, +\infty)$, область изменения функций $[0, +\infty)$. Функция возрастает на всей области определения. Графиками функций являются параболы, ветви которых направлены по оси Oy при $\alpha > 1$ и по оси Ox при $\alpha < 1$ (см.рис.2.9а)

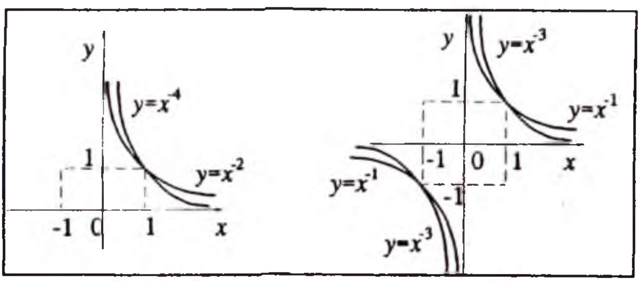


Рис. 2.8а *Рис. 2.8б*

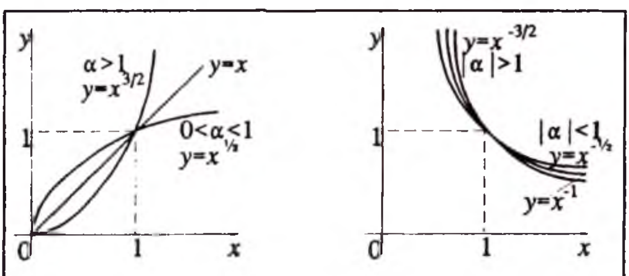


Рис. 2.9а *Рис. 2.9б*

$\alpha < 0$. Область определения бесконечный интервал $(0, +\infty)$, область изменения $(0, +\infty)$. Функция имеет асимптоты $x = 0$, $y = 0$, убывает на всей области определения. Графиком этих функций является одна ветвь гиперболы, расположенная в первой четверти (см.рис.2.9б).

2.3. Построение графиков сложных функций методом преобразования графиков

Для построения графиков функция имеется несколько способов, одним из которых является преобразование графика подходящей элементарной функции до его совпадения с графиком данной функции. Этот способ особенно удобен, если данная функция может быть выражена через одну из элементарных функций в виде $f(x) = k_1 f_{элем}\{k_2(x-a)\} + b$. Тогда график этой функции будет представлять комбинацию сдвигов (горизонтальных и вертикальных), а также сжатий-растяжений графика соответствующей элементарной функции. Рассмотрим более подробно приемы построения графиков функций через их преобразования.

Основные преобразования графика функции

1. Вертикальный сдвиг $f(x) \to f(x) + b$.
2. Горизонтальный сдвиг $f(x) \to f(x+a)$.
3. Комбинированный сдвиг $f(x) \to f(x+a) + b$.
4. Отражение:

а) функции $(f(x) \to -f(x))$, б) аргумента $(f(x) \to f(-x))$.

5. Растяжение $(k > 1)$ / Сжатие $(0 < k < 1)$:

а) функции $(f(x) \to kf(x))$, б) аргумента $(f(x) \to f(kx))$.

6. Взятие модуля:

а) функции $(f(x) \to |f(x)|)$, б) аргумента $(f(x) \to f(|x|))$.

1) Построить график $y = f(x) + b$, если дан график $y = f(x)$. Если $b > 0$, то каждая ордината увеличится на величину b и график "сдвинется вверх" вдоль Oy. Если $b < 0$, то график "сдвинется вниз" на величину b вдоль Oy (см. рис.2.10).

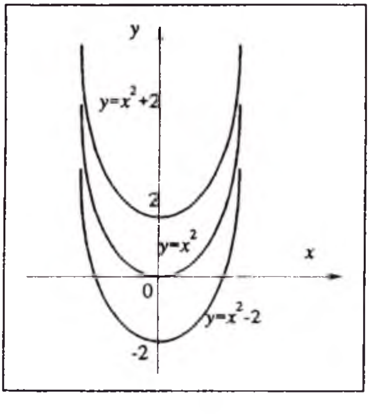


Рис. 10

2) Построить график $y = f(x+a)$, если задан график $y = f(x)$. Имеем $f(x^0)=f(x^1+a)=y^0$, откуда следует, что $x^0=x^1+a$ и следовательно $x^1=x^0-a$. Если $a>0$, то точка x^0 переходит в точку x^1 путём сдвига *влево* на a единиц. Если $a<0$, то точка x^0 переходит в точку x^1 путём сдвига *вправо* на $|a|$ единиц (см. рис. 2.11).

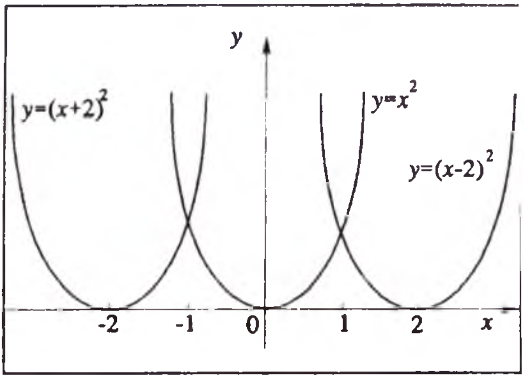


Рис. 2.11

3) Дан график функции $y = f(x)$ (на рис. 2.12 он обозначен пунктиром). Построить график $y = f(x+a) + b$. Строим новую систему координат $x'O'y'$ с началом в точке O' с координатами $(-a;b)$ и осями $O'x' \parallel Ox$, $O'y' \parallel Oy$. Относительно этих осей строим график $y = f(x)$. Полученный график будет искомый в системе координат xOy (см. рис. 2.12).

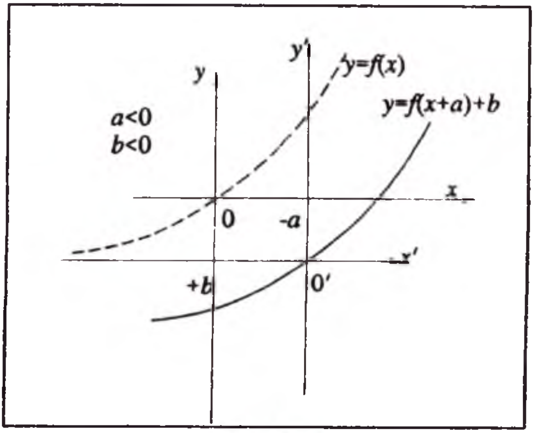


Рис. 2.12

4а) Построить график $y = -f(x)$, если дан график $y = f(x)$. Каждому значению независимой переменной x будет соответствовать противоположное по знаку значение ординаты y. Следовательно, график новой функции будет зеркальным отражением прежней функции относительно оси Ox (см. рис. 2.13а).

4б) Построить график $y = f(-x)$, если задан график $y = f(x)$. Каждому значению ординаты y будет соответствовать противоположное по знаку значение независимой переменной x. Следовательно, график новой функции будет зеркальным отражением прежней функции относительно оси Oy (см. рис. 2.13б).

5а) Построить график $y = kf(x)$, если дан график $y = f(x)$. Если $k > 1$, то каждая ордината увеличится в k раз и график "растянется" вдоль Oy. Если $k < 1$, то график "сожмется" в k раз вдоль Oy (см. рис. 2.14а).

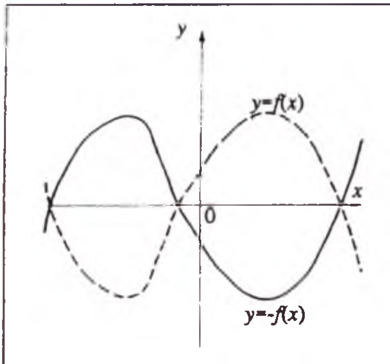


Рис. 2.13а

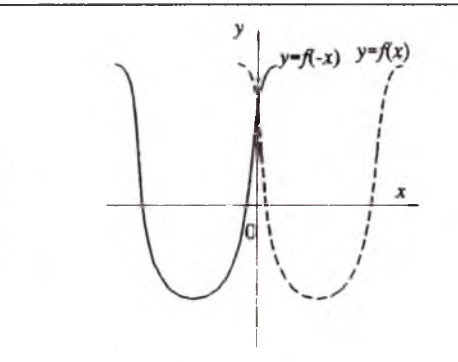


Рис. 2.13б

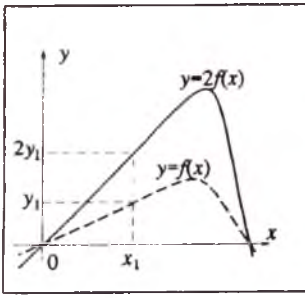


Рис. 2.14а

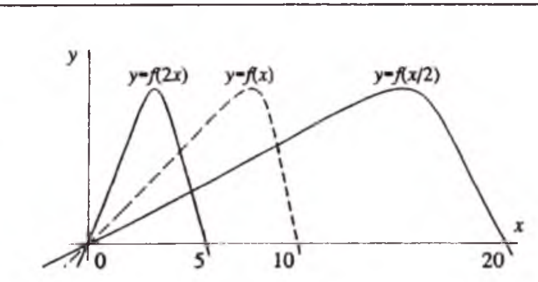


Рис. 2.14б

5б) Построить график $y = f(kx)$, если задан график $y = f(x)$. Если $k > 1$, график "сожмется" в k раз вдоль оси Ox. Если $k < 1$, то график "растянется" в $\frac{1}{k}$ раз вдоль оси Ox (см. рис. 2.14б).

6а) Пусть дан график функции $y = f(x)$ (пунктирный). Надо построить график $y = |f(x)|$. Если $f(x) \geq 0$, то $|f(x)| = f(x)$, следовательно, часть графика функции $y = f(x)$, расположенную в верхней полуплоскости, надо оставить без изменения. Если $f(x) \leq 0$, то $|f(x)| = -f(x)$, значит часть графика, расположенную в нижней полуплоскости, надо отобразить в верхнюю полуплоскость симметрично относительно Ox (см. рис. 2.15а).

6б) Пусть дан график функции $y = f(x)$ (пунктирный). Построить $y = f(|x|)$. Эта функция четная, поэтому график ее симметричен относительно Oy. Но при $x \geq 0$ $f(|x|) = f(x)$, т.е. при $x \geq 0$ графики $f(|x|)$ и $f(x)$ совпадают. Итак, для построения графика нужно часть графика, расположенную в правой полуплоскости, отобразить в левую полуплоскость симметрично относительно оси Oy (см. рис.2.15б).

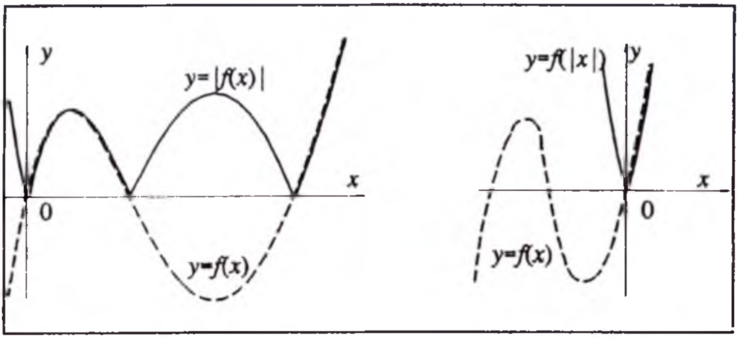


Рис. 2.15а Рис. 2.15б

2.4. Примеры построения и анализа графиков функций одной переменной: квадратный трехчлен, многочлен, дробно-линейные и дробно-рациональные функции

Пример 1. Квадратичная функция $y = ax^2 + bx + c$ $(a \neq 0)$. Функция определена на всей оси Ox. Графиком является парабола, вершина которой имеет координаты $\left(-\frac{b}{2a}, -\frac{b^2 - 4ac}{4a}\right)$. Это легко доказать, выделяя в функции полный квадрат с помощью тождественного преобразования:

$$y = ax^2 + bx + c \equiv a\left[x^2 + 2\left(\frac{b}{2a}\right)x + \left(\frac{b}{2a}\right)^2\right] - \frac{b^2}{4a} + c \equiv$$

$$a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c.$$

Вспоминая преобразования графиков, мы видим, что квадратичная функция представляет собой параболу, смещенную вдоль оси x на величину $-\frac{b}{2a}$ и вдоль оси y на величину $c - \frac{b^2}{4a} \equiv -\frac{b^2 - 4ac}{4a}$. При $a > 0$ ветви параболы направлены вверх, при $a < 0$ - вниз (см. рис.16). Если $a > 0$, то ОИФ $\left[-\frac{b^2 - 4ac}{4a}, +\infty\right)$, при $a < 0$ ОИФ $\left(-\infty, -\frac{b^2 - 4ac}{4a}\right]$. При $b \neq 0$ функция ни четная, ни нечетная. Если дискриминант $D = b^2$-$4ac$ больше нуля, то функция имеет два нуля x_1 и x_2 $\left(x_1 = -\frac{b}{2a} - \frac{\sqrt{D}}{2a} \text{ и } x_2 = -\frac{b}{2a} + \frac{\sqrt{D}}{2a} \right)$, причем, если $a > 0$, то функция положительна при $x < x_1$ и $x > x_2$ и отрицательна в интервале $x_1 < x < x_2$. Если $D = 0$, то функция имеет единственный нуль в точке $-\frac{b}{2a}$. При всех других значениях x она положительна, если $a > 0$ и отрицательна, если $a < 0$. Если $D < 0$, то функция не имеет нулей и все ее значения имеют знак коэффициента a.

При $a > 0$ функция монотонно убывает при $x < -\frac{b}{2a}$ и монотонно возрастает при $x > -\frac{b}{2a}$; при $a < 0$ функция возрастает при $x < -\frac{b}{2a}$ и монотонно убывает при $x > -\frac{b}{2a}$ (см. рис. 2.16).

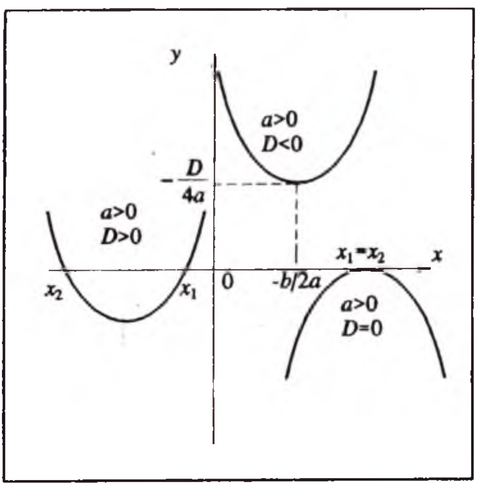


Рис. 2.16

Пример 2. Многочлен $y = (x+2)(x+1)x(x-1)(x-2) \quad x^5 - 5x^3 + 4x$. Этот многочлен представляет собой полином пятой степени и, в соответствии с теорией, имеет не более 5 действительных корней. В данном примере многочлен имеет ровно пять корней: $x_1=-2$, $x_2=-1$, $x_3=0$, $x_4=+1$, $x_5=+2$, т.е. пять точек пересечения графиком оси абсцисс. Данный многочлен представляет собой нечетную функцию аргумента x, определенную на всей числовой оси, график которой показан на рис. 2.16.

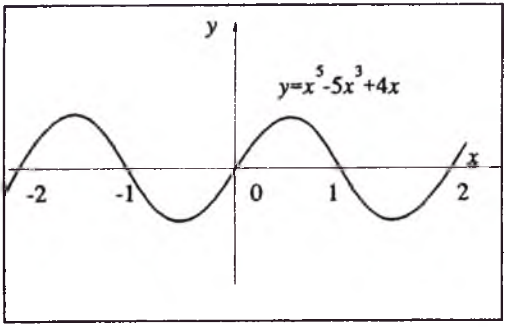


Рис. 2.17

Пример 3. Дробно-линейная функция $y = \dfrac{ax+b}{cx+d}$. Функция определена при $x \neq -\dfrac{c}{d}$. ОИФ (-∞, +∞). Преобразуем выражение, задающее ее, выделением целой части

$$y = \frac{\frac{a}{c}cx + b}{cx+d} = \frac{\frac{a}{c}(cx+d) - \frac{ad}{c} + b}{cx+d} = \frac{a}{c} + \frac{b - \frac{ad}{c}}{cx+d} = \frac{a}{c} + \frac{\frac{bc-ad}{c^2}}{x + \frac{d}{c}}.$$

Введем обозначения $y_0 = \dfrac{a}{c}$; $x_0 = -\dfrac{d}{c}$; $k = \dfrac{bc-ad}{c^2}$. В этих обозначениях можно записать дробно-линейную функцию в виде $y = y_0 + \dfrac{k}{x - x_0}$.

Из такой формы записи сразу видно, что графиком дробно-линейной функции является гипербола, центр которой смещен в точку (x_0, y_0), т.е. в точку $\left(-\dfrac{d}{c}, \dfrac{a}{c}\right)$. Если у гиперболы $y = \dfrac{1}{x}$ асимптотами были оси координат, то у дробно-линейной функции асимптотами будут прямые $x = -\dfrac{d}{c}$; $y = \dfrac{a}{c}$. Знак числа k указывает, в какой четверти относительно асимптот будет лежать гипербола и во сколько раз график надо сжать или растянуть.

В качестве примера построим график функции $y = \dfrac{2x+4}{x+3}$.

1. Выделим целую часть. Разделив двучлены столбиком, получим $y = 2 - \dfrac{2}{x+3}$.

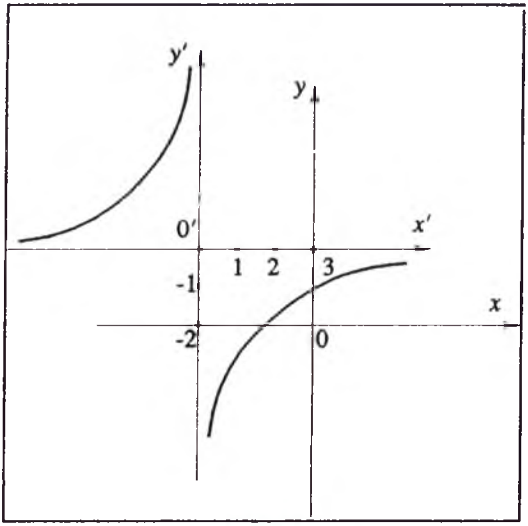


Рис. 2.18

2. На вспомогательной координатной плоскости $x'Oy'$ строим график функции $y' = \frac{2}{x'}$ (см. рис. 2.18). Он расположен во второй и четвертой четвертях.

3. Построим нужную систему координат xOy, сместив вспомогательную систему $x'Oy'$ на 3 единицы вправо и на 2 единицы вниз, осуществив параллельный перенос. В системе xOy, построенной ранее, график будет графиком заданной функции $y = \frac{2x + 4}{x + 3}$ причем действительно прямые $x = -3$ и $y = 2$ являются асимптотами.

В качестве экономического примера дробно-линейной зависимости рассмотрим одну из функций Торнквиста $x(I) = \frac{\alpha I}{I + \beta}$. Эта функция описывает зависимость величины спроса на предметы первой необходимости от величины дохода. Добавляя β к величине дохода в числителе и вычитая β, преобразуем эту функцию к виду: $\frac{\alpha I}{I + \beta} = \frac{\alpha(I + \beta - \beta)}{1 + \beta} = \alpha - \frac{\alpha\beta}{I + \beta}$. Таким образом, задача построения графика этой функции сводится к преобразованию графика гиперболы $\frac{1}{I}$ путем ее растяжения в αβ раз, инвертирования отражения относительно горизонтальной оси и сдвига на величину α вдоль оси x и на величину -β вдоль оси I.

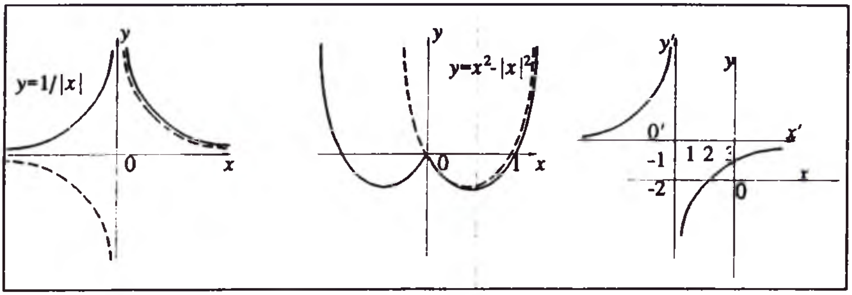


Рис. 2.19а Рис. 2.19б Рис. 2.19в

Пример 4. Построить $y = \frac{1}{|x|}$. Строим $y = \frac{1}{x}$ (пунктирно) и нижнюю ветвь отражаем симметрично *Ox* вверх (изображено на рис.19а сплошной линией).

Пример 5. Построить $y = x^2 - |x|$. Строим $y = x^2 - x$ (пунктирно) и правую часть отражаем симметрично *Oy* влево (см. рис.19б)

Пример 6. Строим график $y = \frac{2x-1}{x+3}$. Выделяя целую часть, получаем $y = 2 - \frac{7}{x+3}$. Строим вспомогательные оси $x = -3$, $y = 2$. Относительно вспомогательных осей строим график $-\frac{7}{x}$ (см. рис. 2.19в).

2.5. Графики в экономическом моделировании

а) Функция потребления и линия бюджетного ограничения;

В теории потребительского спроса на два блага *x* и *y* (к примеру, исследуемое *x* и все остальные *y*) предпочтения потребителя описываются кривой безразличия $U(x,y) = U_k$, а бюджетное ограничение (расходы потребителя $\le$ его дохода) в случае, когда потребитель тратит весь свой доход на рассматриваемые блага: $xp_x + yp_y = I$, где I - доход потребителя, а p_x и p_y - цены благ *x* и *y* соответственно. Для того, чтобы построить графики этих неявно заданных функций $y(x)$ в системе координат, где по оси абсцисс отложена величина блага *x*, а по оси ординат - *y*, нужно выразить в явном виде величину *y* как функцию *x* для обеих зависимостей. Сделаем это для простейшей функции полезности $U(x,y)=xy$. Для уровня полезности (благосостояния) U_0 и дохода I получаем следующие функции: $y = \frac{U_0}{x}$, $y = \frac{I}{p_y} - \frac{p_x}{p_y}x$. Графиком первой из этих функций (кривой безразличия) является гипербола, а графиком второй (бюджетного ограничения) - прямая линия, имеющая отрицательный наклон, равный по абсолютной величине относительной цене блага x и точку пересечения с осью ординат $\left(\frac{I}{p_y}\right)$ соответствующую количеству блага y, которое можно приобрести по цене p_y, если потратить на него весь доход I (см. рис. 2.20).

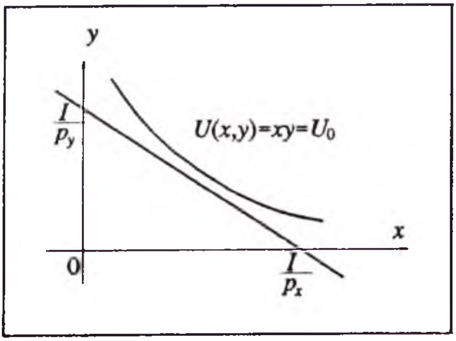


Рис. 2.20

б) Кривые спроса и предложения.

Другим примером функций в экономике служат функции спроса и предложения $p(q)$, выражающие связь цены блага и величины спроса или предложения блага при постоянных вкусах потребителей, ценах на другие блага и других параметрах. Пример графика линейной функции спроса приводился в самом начале главы. Аналогично строится и график функции предложения, но в отличие от функции спроса он отражает положительную связь переменных p и q (см. рис. 2.21).

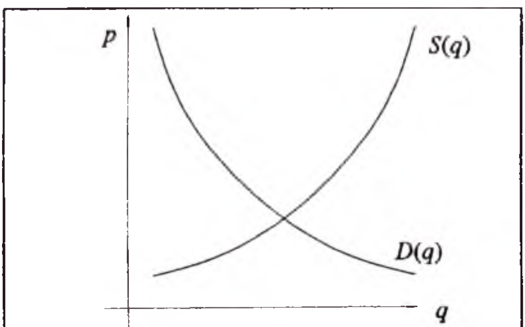


Рис. 2.21

в) Зависимости величины спроса от дохода.

В модели потребительского спроса используются также функции Торнквиста, моделирующие связь между величиной дохода (I) и величиной спроса потребителей (x) на: *а)* малоценные товары $\left(x = \frac{\alpha I(I+\beta)}{I^2+\gamma}\right)$; *б)* товары первой необходимости $\left(x = \frac{\alpha I}{I+\beta}\right)$; *в)* товары второй необходимости (относительной роскоши) $\left(x = \frac{\alpha(I-\gamma)}{I+\beta}\right)$;

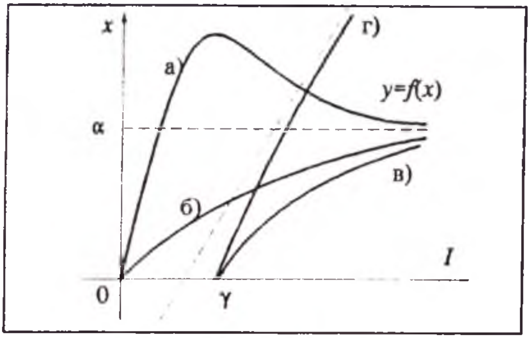


Рис. 2.22

г) предметы роскоши $\left| x = \dfrac{\alpha I(I - \gamma)}{I + \beta} \right|$. Соответствующие им графики приведены на рис. 2.22.

г) Графики зависимости издержек и дохода от объема производства.

В качестве последнего примера рассмотрим функции издержек $C(q)$ и дохода фирмы $R(q) = qp(q)$ в зависимости от объема производства q. Поведение функции дохода определяется функцией спроса $p(q)$, рассмотренной выше. Поэтому рассмотрим более подробно поведение функции издержек. В типичном случае издержки фирмы велики при небольшом объеме производства q и вначале растут быстрее, чем доход. С увеличением объема производства скорость роста издержек уменьшается и в какой-то момент они сравниваются с доходом и фирма начинает получать прибыль. При увеличении объема производства прибыль увеличивается, достигая максимума при оптимальном значении q. При дальнейшем увеличении объема производства издержки снова начинают расти быстрее дохода (исчерпанны эффективные ресурсы, нужны дополнительные помещения, сырье, квалифицированная рабочая сила) и прибыль фирмы уменьшается, достигая отрицательных значений при достаточно больших объемах производства. Типичные графики дохода издержек и прибыли приведены на рис. 2.23. Им, например, могут соответствовать функции $R(q) = aq - bq^2$, $C(q) = cq - dq^2 + eq^3$.

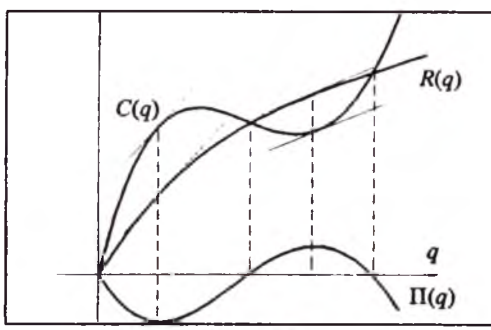


Рис. 2.23

Вопросы к главе 2

1. Приведите определение функции.
2. Что такое область определения и область допустимых значений функции?
3. Какие существуют способы задания функций? Приведите конкретные примеры каждого способа.
4. Дайте определения возрастания и убывания функции. Приведите примеры возрастающей и убывающей функций.
5. Какие функции называются четными (нечетными)?
6. Дайте определение асимптоты. Приведите пример функции, имеющей горизонтальную асимптоту. Приведите пример функции, имеющей вертикальную асимптоту.
7. Приведите пример линейных функций, описывающих зависимости спроса и предложения от цены товара. Постройте их графики.
8. Приведите пример степенной функции и постройте ее график.
9. Приведите пример показательных функций, описывающих зависимости спроса и предложения от цены товара. Постройте их графики.
10. Приведите пример логарифмической функции и постройте ее график.
11. Постройте графики функций:

$$y=2x;\ y=0{,}5x;\ y=x+2;\ y=x-2;\ y=2(x-2);\ y=\frac{1}{2(x+2)};$$

$$y=2(x+2)+2;\ y=\frac{1}{2(x-2)-2}$$

методом преобразования графиков.
12. Постройте кривые безразличия функции полезности $y=x_1 x_2$ при уровнях полезности, равных 2 и 3. Найдите их асимптоты.
13. Приведите пример функции, описывающей бюджетное ограничение. Найдите ее точки пересечения с осями координат.
14. Приведите пример функции, описывающей зависимость величины спроса от дохода.
15. Приведите пример функции, описывающей зависимость предложения от цены. Постройте ее график.

ГЛАВА 3
ОСНОВЫ ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ. ДИФФЕРЕНЦИАЛЬНОЕ ИСЧИСЛЕНИЕ В ЭКОНОМИЧЕСКОМ АНАЛИЗЕ

3.1. Экономические задачи, решаемые методами дифференциального исчисления

Дифференциальное исчисление - широко применяемый для экономического анализа математический аппарат. Базовой задачей экономического анализа является изучение связей экономических величин, записываемых в виде функций. В каком направлении изменится доход государства при увеличении налогов или при введении импортных пошлин? Увеличится или уменьшится выручка фирмы при повышении цены на ее продукцию? В какой пропорции дополнительное оборудование может заменить выбывающих работников? Для решения подобных задач должны быть построены функции связи входящих в них переменных, которые затем изучаются с помощью методов дифференциального исчисления.

В экономике очень часто требуется найти наилучшее, или оптимальное значение того или иного показателя: наивысшую производительность труда, максимальную прибыль, максимальный выпуск, минимальные издержки и т.д. Каждый показатель представляет собой функцию одного или нескольких аргументов. Например, выпуск можно рассматривать как функцию затрат труда и капитала (как это делается в производственных функциях). Таким образом, нахождение оптимального значения показателя сводится к нахождению экстремума (максимума или минимума) функции одной или нескольких переменных. Подобные задачи порождают класс экстремальных задач в экономике, решение которых требует использования методов дифференциального исчисления. Если экономический показатель y нужно максимизировать или минимизировать как функцию другого показателя x (например, задача на максимум прибыли как функции объема выпуска), то в оптимальной точке (т.е. в точке максимума) приращение функции y на приращение аргумента x должно стремиться к нулю, когда приращение аргумента стремится к нулю. Иначе, если такое приращение стремится к некоторой положительной или отрицательной величине, рассматриваемая точка не является оптимальной, поскольку увеличив или уменьшив аргумент x, можно изменить величину y в нужном направлении. В терминах дифференциального исчисления это означает, что необходимым условием экстремума функции $y=f(x)$ является равенство нулю ее производной.

В экономике часто приходится решать задачи на экстремум функций нескольких переменных, поскольку экономические показатели обычно зависят от многих факторов. Такие задачи хорошо изучены теорией функций нескольких переменных, использующей методы дифференциального исчисления. Многие задачи включают не только максимизируемую (минимизируемую) функцию, но и ограничения (скажем, бюджетное ограничение в задаче потребительского выбора). Это - задачи математического программирования, для решения которых разработаны специальные методы, также опирающиеся на дифференциальное исчисление. Все эти виды задач и их приложения будут рассмотрены в последующих главах; мы не будем здесь забегать вперед.

Важный раздел методов дифференциального исчисления, используемых в экономике, называется методами предельного анализа. Предельный анализ в экономике - совокупность приемов исследования изменяющихся величин затрат или результатов при изменениях объемов производства, потребления и т.п. на основе анализа их предельных значений. Предельный показатель (показатели) функции $y=f(x)$ - это ее производная (в случае функции одной переменной) или частные производные (в случае функции нескольких переменных).

В экономике широко используются средние величины: средняя производительность труда, средние издержки, средний доход, средняя прибыль и т.д. Но часто требуется узнать, на какую величину вырастет результат, если будут увеличены затраты или, наоборот, насколько уменьшится результат, если затраты сократятся. С помощью средних величин ответ на этот вопрос получить невозможно. В подобных задачах требуется определить предел отношения приростов результата и затрат, т.е. найти предельный эффект. Следовательно, для их решения необходимо применение методов дифференциального исчисления - нахождение производной в случае функции одной переменной и частных производных, если функция зависит от нескольких аргументов.

Так, например, если задана производственная функция:

$$y = f(x_1, \ldots, x_i, \ldots, x_n),$$

где x_i - объём затрачиваемого i-го ресурса ($i = 1,\ldots,n$), y - максимальный объем выпуска, который можно получить, затрачивая ресурсы соответственно в объёмах $x_1,\ldots,x_i,\ldots,x_n$, то предельный эффект от использования i-го ресурса (p_i) определяется следующим образом:

$$p_i = \frac{f(x_1,\ldots,x_i+\Delta x_i,\ldots,x_n)-f(x_1,\ldots,x_i,\ldots,x_n)}{\Delta x_i}.$$

Здесь величина p_i равна дополнительному объему выпуска, который получается в результате затраты дополнительной единицы Дx_i i-го ресурса при неизменных объёмах остальных ресурсов.

Показатель предельного эффекта в оптимизационных моделях применяется для нахождения оптимального объема производства при заданных ресурсах, а также для определения оптимального распределения ограниченных ресурсов по различным направлениям их использования. Если максимизируемый показатель (например, прибыль) есть разность результата и издержек (в данном случае результат представлен выручкой), то в оптимальной точке предельная выручка должна равняться предельным издержкам. Такое равенство должно выполняться по каждому из факторов, определяющих выручку и издержки, что вытекает из необходимости равенства нулю частных производных прибыли по всем этим факторам. Необходимые и достаточные условия оптимума во многих экономических задачах записываются с помощью частных производных и дифференциалов. Так, если решается задача на максимум выпуска, описываемого с помощью приведенной выше производственной функции, при наличии ограничения по общему расходу денежных средств на используемые в производстве ресурсы, то в оптимальной точке должны быть равны между собой отношения предельных производительностей ресурсов p_i и их цен. Иными словами, для всех ресурсов должен быть одинаков предельный эффект в расчете на единицу дополнительно расходуемых на эти ресурсы денежных средств. В задаче потребительского выбора отношение предельных полезностей благ должно быть равно отношению их цен. Иначе говоря, предельная полезность в расчете на одну денежную единицу должна быть в оптимальной точке одинакова по всем благам; в противном случае бюджет потребителя мог бы быть перераспределен с увеличением его благосостояния. Таким образом, методы дифференциального исчисления позволяют не только решить различные экономические задачи, но и записать необходимые или достаточные условия оптимума в этих задачах, которые позволяют дать ответ на те или иные конкретные вопросы.

Широко используется в экономическом анализе понятие дифференциала, или главной линейной части приращения функции. Так, если некоторая величина y есть функция двух аргументов x_1 и x_2, то с использованием дифференциала легко рассчитать предельную норму замены между этими аргументами, т.е. величину, показывающую, сколько нужно фактора 2 для замены одной единицы фактора 1 с сохранением значения функции y. Предельная норма замены важна в задачах потребительского выбора (взаимозаменяемость благ), в задачах оптимизации производства (взаимозаменяемость труда и капитала) и в ряде других задач. Пусть $y=f(x_1,x_2)$. Если мы хотим сохранить значение функции y неизменным, то это означает, что приращение y, а значит и его главная линейная часть должны быть равны нулю. Иными словами, $0 = dy = y'_{x1}dx_1 + y'_{x2}dx_2$. Отсюда предельная норма замены $-\frac{dx_1}{dx_2}=\frac{y'_{x_2}}{y'_{x_1}}$, то есть равняется отношению частных производных функции y по первому и второму факторам.

Методы дифференциального исчисления широко применяются не только для анализа взаимодействия отдельных экономических факторов, определения их взаимозаменяемости или оптимального сочетания, но и в сложных моделях экономики, в частности - в моделях экономической динамики. Дифференциальное исчисление - это не только аппарат, позволяющий находить решения таких моделей, но и необходимый составной элемент для их построения. Динамические модели применяются для решения таких задач, как определение оптимальной или равновесной траектории развития экономической системы, ее состояний в заданные моменты времени, анализ системы на устойчивость, анализ структурных сдвигов и т.п. Некоторые модели этого типа будут рассмотрены в главе 12.

Из рассмотренных направлений применения дифференциального исчисления в экономике важнейшим является вопрос нахождения и анализа взаимосвязей экономических переменных, определяющих функционирование экономического объекта или протекание экономического явления, который мы сейчас рассмотрим более подробно.

3.1.1. Анализ взаимосвязей экономических показателей

Анализируя взаимосвязи экономических показателей, мы должны последовательно ответить на четыре вопроса: Какие факторы определяют интересующий нас экономический показатель? Каков знак этой зависимости?

Какова степень этой зависимости? Каково числовое (функциональное) выражение соответствующей зависимости? Рассмотрим возможные ответы на эти вопросы на примере простейшей экономической зависимости - функции спроса.

- **От чего зависит** (от каких факторов)?

В ответ на этот вопрос надо перечислить все факторы, определяющие исследуемый экономический показатель. В частности, величина спроса q^D на какой-либо товар определяется ценой этого товара p, доходом потребителей I, ценами на другие товары (дополняющие (C) или заменяющие (S) данный товар), ожидаемыми ценами и ожидаемым доходом. Сокращенно это можно записать так:

$$q^D = f(p, I, T, p^C, p^S, p^e, I^e, \ldots).$$

- Как зависит (положительно или отрицательно)?

В ответ на этот вопрос надо определить характер взаимосвязи. Исследуемый показатель связан с каким-либо фактором положительно, если его значение возрастает при увеличении фактора и отрицательно, если его значение уменьшается при увеличении фактора. В частности, величина спроса q^D на какой-либо товар уменьшается при увеличении его цены p, увеличивается (для нормальных товаров) или уменьшается (для некачественных товаров) при увеличении дохода потребителей I, уменьшается при увеличении цен на дополняющие товары и увеличивается при увеличении цен на заменяющие данный товар товары, увеличивается при ожидании повышения цен или доходов. Сокращенно это можно записать так:

$$q^D = f(\overset{-}{p}, \overset{\pm}{I}, \overset{+}{T}, \overset{-}{p^c}, \overset{+}{p^s}, \overset{+}{p^e}, \overset{\pm}{I^e}, \ldots).$$

- Какова степень зависимости?

Для ответа на этот вопрос надо определить насколько чувствителен исследуемый экономический показатель к изменению определяющих его факторов? Другими словами, какова степень его изменения при заданном абсолютном или относительном изменении факторов.

$$\Delta p = \in, \ \Delta q^D = ? \text{ или } \Delta I = \in, \ \Delta q^D = ?$$

Имеются два подхода к анализу чувствительности зависимости $y = f(x)$.

- Приростной подход ($\Delta x \Rightarrow \Delta y$).

прирост фактора $\Rightarrow$ прирост исследуемого показателя

(изменение x) $\Rightarrow$ (изменение y)

Мера “абсолютной” чувствительности - скорость изменения функции (средняя (отношение изменений) или предельная (производная)):

$$\frac{\Delta y}{\Delta x} \to \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx} = f'(x).$$

- Темповый подход ($\%\Delta x \Rightarrow \%\Delta y$).

темп прироста фактора $\Rightarrow$ темп прироста исследуемого показателя

(процентное изменение x) $\Rightarrow$ (процентное изменение y)

Напомним, что процентное изменение какой-либо переменной - это отношение изменения этой переменной к первоначальному ее значению:

$$\%\Delta x = \frac{\Delta x}{x} = \frac{x_2 - x_1}{x_1}.$$

Например, если цена буханки хлеба увеличилась с 200 до 300 рублей, то процентное изменение цены

$$\%\Delta p = \frac{300-200}{200} = 50\%.$$

Мера "относительной" чувствительности - эластичность функции (средняя (отношение процентных изменений) или предельная ($\approx$ производной)):

$$\frac{\%\Delta y}{\%\Delta x} \to \lim_{\%\Delta x \to 0} \frac{\%\Delta y}{\%\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta y / y}{\Delta x / x} = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{f'(x) \cdot x}{y}.$$

- Каково функциональное выражение зависимости?

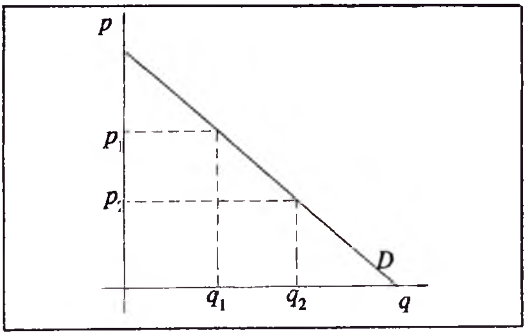
Для ответа на этот вопрос надо указать конкретное функциональное выражение исследуемой зависимости (в виде формулы, графика или таблицы). Эту зависимость можно получить либо из теоретической модели или из эконометрического (эмпирического) исследования. Например, функция спроса на какой-либо товар может определяться следующим выражением:

$$q^D = q_0 - \alpha \cdot p - \beta \cdot p^C + \gamma \cdot p^S$$

или изображаться прямой линией на графике.

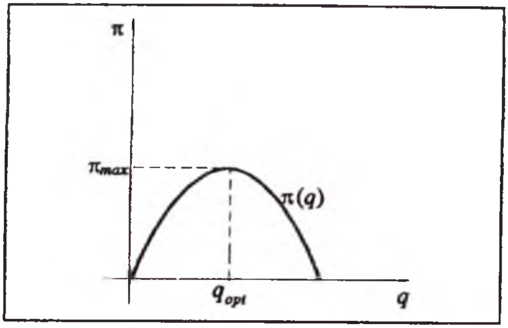
Другим не менее важным направлением дифференциального исчисления является его применение к принятию оптимальных решений, которое мы также рассмотрим более подробно, чем во введении.

3.1.2. Принятие оптимальных решений



Пусть, например, монополист, зная (из маркетинговых исследований) функцию спроса на свой товар, решает, сколько ему производить и по какой цене продавать свой товар.

Если он установит достаточно высокую цену, то потребители за определенный период купят у него не слиш-



ком много товара. Если он будет производить больше, то ему придется понизить цену, чтобы распродать все выпускаемое им количество за определенный период времени. При этом выручка увеличится за счет увеличения объема продаж (выигрыш) и одновременно уменьшится за счет уменьшения цены (потери). Результат будет зависеть от того, что окажется большим: выигрыш или потери. Как же все-таки монополист может определить оптимальный объем выпуска? Для этого он должен определить зависимость выручки (или прибыли, если учитывать издержки выпуска) от объема выпуска $p(q) = R(q)-C(q) = p(q)\, q - C(q)$ и определить, при каком объеме выпуска прибыль будет максимальна.

Из теории известно, что задача определения максимума функции решается с помощью понятия производной. Для этого надо знать ответы на два вопроса.

1. Как находить производные произвольных функций?
2. Как применять производные к исследованию функций?

Для ответа на первый вопрос мы рассмотрим определение и геометрический смысл производной, формулы для нахождения производных нескольких простейших (элементарных) функций и правила дифференцирования, позволяющие находить производные от любых комбинаций элементарных функций.

Для ответа на второй вопрос мы рассмотрим связь знака и величины производной с возрастанием, убыванием функций и определим необходимые и достаточные условия экстремума (максимума или минимума) функций.

3.2. Приращение величины, аргумента, функции. Скорость изменения функции

Приращение величины, аргумента, функции. Пусть величина z меняется от значения z_1 (начальное значение) до значения z_2 (конечное значение). Тогда величина $\Delta z = z_2 - z_1 = z_{кон} - z_{нач}$ называется приращением величины z. Приращение возрастающей величины ($z_{кон} > z_{нач}$) будет положительно $\Delta z > 0$, а приращение убывающей величины ($z_{нач} > z_{кон}$) будет отрицательно $\Delta z < 0$. Если величина z не изменилась ($z_{нач} = z_{кон}$), то ее приращение будет равно нулю $\Delta z = 0$.

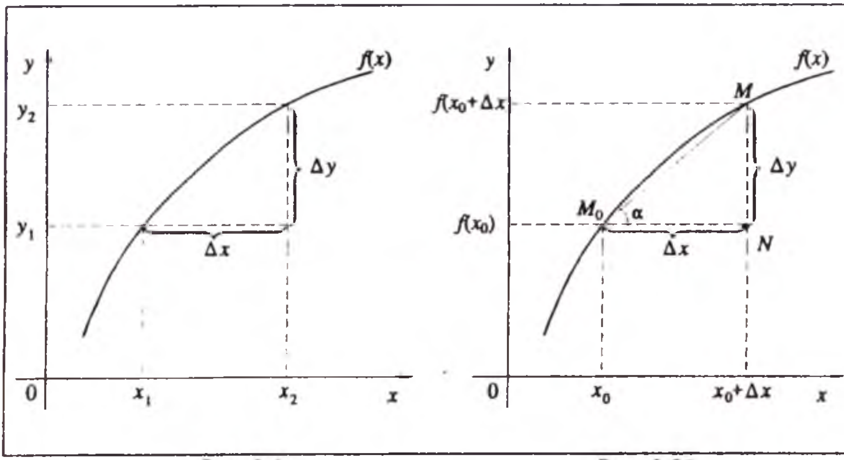


Рис. 3.1а *Рис. 3.2б*

Пусть дана функция $y = f(x)$ и два значения аргумента x_1 и x_2. Им соответствуют два значения функции $y_1 = f(x_1)$ и $y_2 = f(x_2)$. Разность $\Delta x = x_2 - x_1$ называется приращением аргумента, а $\Delta y = y_2 - y_1 = \Delta f = f(x_2) - f(x_1)$ - приращением функции. Геометрическая интерпретация этих величин показана на рис. 3.1а.

Рассмотрим функцию $f(x)$, x $[a;b]$. Возьмем произвольную точку x_0 $[a;b]$. Для любого x $[a;b]$ разность $x - x_0$ называется приращением аргумента x в точке x_0 и обозначается Δx. Таким образом, $\Delta x = x - x_0$, $x = x_0 + \Delta x$. Разность $f(x) - f(x_0)$ называется приращением функции $f(x)$ в точке x_0 и обозначается $\Delta f(x_0)$ (или Δf, или Δy) (см. рис. 3.1б). Следовательно,

$$\Delta f(x_0) = f(x) - f(x_0) = f(x_0 + \Delta x) - f(x_0).$$

В этих терминах можно сказать, что функция $f(x)$ непрерывна в точке x_0 тогда и только тогда, когда приращение функции в точке x_0 стремится к нулю, если приращение аргумента стремится к нулю.

Скорость изменения функции на интервале (средняя скорость).

Рассмотрим две функции, показанные на рис. 3.2. Значения каждой из них меняются при изменении аргумента на величину $\Delta x = x - x_0$. Из рисунка видно, что вторая функция меняется (возрастает) сильнее, чем первая на интервале $(x_0;x)$.

Для сравнения величин изменения различных функций при одинаковом изменении аргумента вводится понятие скорости (быстроты) изменения функции на интервале $(x_0;x)$ (средней скорости), определяемой, как отношение изменения функции, вызванного изменением ее аргумента, к соответствующему изменению аргумента.

$$\frac{\textit{скорость изменения функции на интервале } (x_0;x)}{} = \frac{\textit{изменение функции}}{\textit{изменение аргумента}}.$$

Обозначая скорость буквой v, запишем это соотношение в виде.

$$v_{f(x)}(x_0;\Delta x)=\frac{\Delta f(x_0)}{\Delta x}=\frac{f(x+\Delta x)-f(x_0)}{\Delta x}=\frac{f(x)-f(x_0)}{x-x_0}.$$

Из этого соотношения вытекает, что если при равных изменениях аргумента ($\Delta x^{(1)}=\Delta x^{(2)}$) одна из функций меняется сильнее ($\Delta y^{(2)}>\Delta y^{(1)}$), то и скорость изменения второй функции на рассматриваемом интервале будет больше, чем первой ($v_y(2)>v_y(1)$).

Геометрический смысл скорости изменения функции на интервале $(x_0;x)$ (средней скорости) в том, что она численно равна тангенсу угла наклона отрезка, соединяющего две точки графика функции, соответствующих значениям аргумента x_0 и x, т.е. тангенсу угла α в треугольнике M_0MN (см. рис. 3.1б).

Недостаток такого определения скорости состоит в том, что эта скорость зависит не только от точки x_0, относительно которой рассматривается изменение аргумента, но и от самой величины изменения аргумента, т.е. от величины интервала Δx, на котором определяется скорость. Для устранения этого недостатка вводится понятие скорости изменения функции в точке (мгновенной скорости).

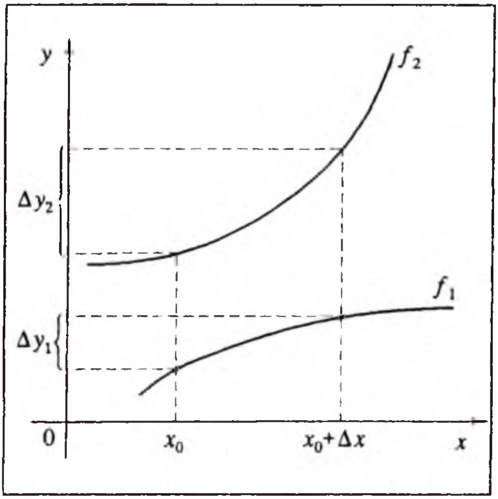


Рис. 3.2

Скорость изменения функции в точке (мгновенная скорость).

Для определения скорости изменения функции в точке x_0 сближают точки x и x_0, устремляя интервал Δx к нулю. Изменение непрерывной функции при этом будет также стремиться к нулю. При этом отношение, стремящегося к нулю изменения функции к стремящемуся к нулю изменению аргумента дает скорость изменения функции в точке x_0 (мгновенной скорости), точнее на бесконечно малом интервале, относительно точки x_0.

$$v_{f(x)}(x_0)=\lim_{\Delta x\to 0}\frac{\Delta f(x_0)}{\Delta x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x_0)}{\Delta x}=\lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}.$$

Именно эту скорость изменения функции $f(x)$ в точке x_0 и называют производной функции $f(x)$ в точке x_0.

Геометрический смысл скорости изменения функции в точке x_0 (мгновенной скорости) в том, что она численно равна тангенсу угла наклона касательной к графику функции в точке x_0. Это непосредственно следует из ее определения, поскольку при сближении точек x_0 и x, точки пересечения графика функции прямой линией M_0 и M также сближаются и сливаются в одной точке M_0, в которой линия и касается графика функции.

3.3. Определение производной и ее геометрический смысл. Правила дифференцирования. Дифференцирование основных элементарных функций: линейной, степенной, показательной и логарифмической функций

Производная функции. Пусть функция $f(x)$ определена в некоторой окрестности точки x_0.

Определение. Производной функции $f(x)$ в точке x_0 называется предел отношения приращения функции $\Delta f(x_0)$ к приращению аргумента Δx при $\Delta x \to 0$, если этот предел существует, и обозначается $f'(x_0)$. Итак,

$$f'(x_0)=\lim_{\Delta x\to 0}\frac{\Delta f(x_0)}{\Delta x}=\lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}.$$

Производную функции $y=f(x)$, $x\in(a;b)$, в точке x обозначают $f'(x)$ (“эф штрих от икс”), y'_x (“игрек штрих по икс”), $\frac{df(x)}{dx}$ (“де эф по де икс”), $\frac{dy}{dx}$ (“де игрек по де икс”), причем все эти обозначения равноправны.

Экономисты используют также символ $Mf(x)$ (т.е. $Mf(x)=f'(x)$) и термин маржинальное значение функции f в точке x.

Из определения производной вытекает следующая схема ее нахождения, которую изложим на конкретном примере.

Пример 1. Найти производную функции $y = f(x) = x^2 - 2x + 3$ в произвольной (но фиксированной) точке x.

1) Даем аргументу x приращение $\Delta x \neq 0$ и находим "приращенное" значение функции: $y + \Delta y = f(x+\Delta x) = (x+\Delta x)^2 - 2(x+\Delta x) + 3 = x^2+2x\Delta x + (\Delta x)^2 - 2x-2\Delta x+3$.

2) Находим приращение функции: $\Delta y = \Delta f(x) = f(x+\Delta x) - f(x) = x^2 + 2x\Delta x + (\Delta x)^2 - 2x - 2\Delta x + 3 - x^2 - 2x + 3 = 2x\Delta x + (\Delta x)^2 - 2\Delta x$.

3) Находим отношение приращения функции к приращению аргумента:

$$\frac{\Delta y}{\Delta x} = \frac{\Delta f(x)}{\Delta x} = \frac{2x\Delta x + (\Delta x)^2 - 2\Delta x}{\Delta x} = 2x + \Delta x - 2.$$

4) Находим предел этого отношения при $\Delta x \to 0$, т.е.искомую производную:

$$y = f(x) = \lim_{\Delta x \to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0}\frac{\Delta f(x)}{\Delta x} = \lim_{\Delta x \to 0}(2x + \Delta x - 2) = 2x - 2 = 2(x-1).$$

Определение. Операция нахождения производной называется дифференцированием функции.

Определение. Функция, имеющая производную в точке x_0, называется дифференцируемой в этой точке.

Функция, имеющая производную в каждой точке интервала $(a;b)$, называется дифференцируемой на этом интервале; при этом производную $f'(x)$ можно рассматривать как функцию на $(a;b)$. Необходимое условие существования производной вытекает из следующего утверждения.

Если функция $f(x)$ дифференцируема в точке x_0, то она непрерывна в этой точке.

Геометрический смысл производной. Пусть $y = f(x)$ - дифференцируемая в точке x_0 функция, график которой изображен на рис.3.3а. M_0T - касательная к графику функции $y = f(x)$ в точке M_0 с абсциссой x_0 и ординатой y_0. Являясь по сути скоростью изменения функции в точке x_0 (точнее в бесконечно малом интервале вблизи точки x_0), производная функции $y = f(x)$ в точке x_0 численно равна тангенсу угла наклона касательной, проведенной к графику этой функции в точке $(x_0;f(x_0))$. Можно показать, что этот вывод не зависит от расположения графика функции и касательной на координатной плоскости (см. рис. 3.3б). Этот вывод следует непосредственно из определения производной функции.

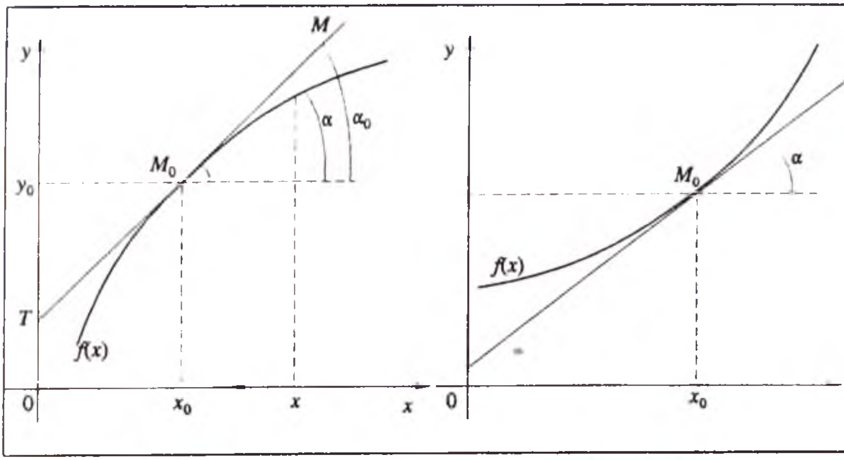


Рис. 3.3а Рис. 3.3б

Правила дифференцирования. Определение производной редко используется для практического вычисления производных функций. Если функция, производную которой нужно найти, представляет из себя комбинацию элементарных функций, то для вычисления производных применяются таблицы производных элементарных функций и правила дифференцирования, важнейшие из которых приведены ниже.

1. Производная суммы (разности) функций равна сумме (разности) производных этих функций: $(u(x) \pm v(x))' = u'(x) \pm v'(x)$.

2. Постоянный множитель выносится за знак производной: $(cu(x))' = cu'(x)$.

3. Правило дифференцирования произведения функций: $(u(x)v(x))' = u'(x)v(x) + u(x)v'(x)$.

4. Правило дифференцирования частного функций:

$$\left[\frac{u(x)}{v(x)}\right]' = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2}.$$

5. Правило дифференцирования сложной функции $(f(u(x)))$: $f'_x(u(x)) = f'_u(u)\, u'_x(x)$
$(y'_x = y'_u\, u'_x)$.

6. Правило дифференцирования обратной функции $(x(y))$: $y'_x = \frac{1}{x'_y}$.

7. Правило дифференцирования неявной функции $0 = F(x,y)$: $y'_x = -\frac{F'_x}{F'_y}$.

Дифференцирование основных элементарных функций: линейной, степенной, показательной и логарифмической функций.

1. $c' = 0$ (производная константы равна нулю).
2. $(a+bx)' = b$ (производная линейной функции равна константе).
3. $(x^{\alpha})' = \alpha x^{\alpha-1}$ $(x > 0)$ (дифференцирование степенной функции уменьшает ее степень на единицу).
4. $(a^x)' = a^x \ln a$ $(0 < a \neq 1)$. В частности $(e^x)' = e^x$.
5. $(\log_a x)' = \frac{1}{x}\log_a e$ $(0 < a \neq 1,\ x>0)$. В частности $(\ln x)' = \frac{1}{x}$.

3.4. Дифференциал функции одной переменной. Приближенные вычисления

Понятие дифференциала функции. Пусть функция $y = f(x)$, x $(a;b)$ дифференцируема в некоторой точке x_0 $(a;b)$, т.е. в точке x_0 существует предел $\lim\limits_{x \to x_0}\frac{\Delta f(x_0)}{\Delta x} = f'(x_0)$. Отсюда следует, что $\frac{\Delta f(x_0)}{\Delta x}$, где α - величина бесконечно малая при $\Delta x \to 0$, т.е. $\lim\limits_{\Delta x \to 0}\alpha = 0$. Из этого равенства находим $\Delta f(x_0) = f'(x_0)\Delta x + \alpha\Delta x$.

Если $f'(x_0) \neq 0$, то слагаемое $f'(x_0)\Delta x$, линейное относительно Δx является бесконечно малым при $\Delta x \to 0$, так как $\lim\limits_{\Delta x \to 0} f'(x_0)\Delta x = 0$.

Второе слагаемое $\alpha\Delta x$ в выражении для приращения функции также является бесконечно малым при $\Delta x \to 0$, потому что $\lim\limits_{\Delta x \to 0}\alpha\Delta x = 0$.

Однако $\lim\limits_{x \to x_0}\frac{\alpha\Delta x}{f'(x_0)\Delta x} = \lim\limits_{x \to x_0}\frac{\alpha}{f'(x_0)} = 0$. Следовательно, слагаемое $\alpha\Delta x$ есть бесконечно малое более высокого порядка, чем слагаемое $f'(x_0)\Delta x$. Поэтому говорят, что величина $f'(x_0)\Delta x$ составляет главную часть приращения функции $f(x)$ в точке x_0.

Определение. Дифференциалом функции $y = f(x)$ в точке x_0 называется линейная относительно Δx величина $f'(x_0)\Delta x$, составляющая главную часть приращения функции $f(x)$ в точке x_0. Дифференциал функции обозначается $df(x_0)$ ("де эф от икс нулевое") или dy ("де игрек"). Таким образом, $df(x_0) = f'(x_0)\Delta x$.

Заметим, что если $f'(x_0) = 0$, то $f'(x_0)\Delta x = 0$, и слагаемое $f'(x_0)\Delta x$ не является главной частью приращения $\Delta f(x_0)$, так как $\alpha\Delta x$, вообще говоря, отлично от нуля. Однако в этом случае по определению полагаем $df(x_0) = f'(x_0)\Delta x = 0$.

Если данная функция дифференцируема в каждой точке интервала $(a;b)$, то пишут $df(x) = f'(x)\Delta x$ или $dy = y'\Delta x$.

Пример 2. Найти дифференциал функции $y = x$. По определению дифференциала имеем $dy = dx = (x')\Delta x = \Delta x$. Итак, дифференциал

dx независимой переменной совпадает с её приращением Δx, т.е. $dx = \Delta x$. Поэтому его определение можно записать в виде $df(x) = f'(x)dx$ или $dy = y'dx$, т.е. дифференциал функции $y = f(x)$ равен произведению производной этой функции на дифференциал ее аргумента.

Найдем дифференциал сложной функции $y = f(u)$, где $u = g(x)$ По определению дифференциала находим $dy = y'_x dx = y'_u u'_x dx$. Но $u'_x dx = du$, поэтому $dy = y'_u du$. Таким образом, форма дифференциала не зависит от того, является аргумент данной функции независимой переменной или функцией другого аргумента. Это свойство дифференциала сложной функции называется инвариантностью формы дифференциала. Следует, однако, заметить, что в последней формуле нельзя заменить du на Δu, так как du Δu для любой функции u, кроме линейной.

Геометрический смысл дифференциала. Пусть $y = f(x)$ - дифференцируемая в точке x_0 функция, график которой изображен на рис. 3.4а, M_0T - касательная к графику функции $y = f(x)$ в точке M_0 с абсциссой x_0. Рассмотрим ординату этой касательной, соответствующую абсциссе $x_0 + \Delta x$. Из прямоугольного треугольника ΔM_0NT находим $NT = M_0N\operatorname{tg}\alpha$, но $M_0N = \Delta x$ и $\operatorname{tg}\alpha = f'(x_0)$. Поэтому $NT = f'(x_0)\Delta x = df(x_0)$. Таким образом, дифференциал функции $y = f(x)$ в точке x_0 равен приращению ординаты касательной, проведенной к графику этой функции в точке $(x_0;f(x_0))$, соответствующему приращению ее абсциссы x_0 на Δx. Можно показать, что этот вывод не зависит от расположения графика функции и касательной на координатной плоскости (см. рис. 3.4б).

Дифференциал может быть как меньше приращения функции (см. рис. 3.4а), так и больше (см. рис. 3.4б). Однако при достаточно малых приращениях Δx можно принять $\Delta f(x_0)$ $df(x_0)$. Этот вывод следует непосредственно из определения дифференциала функции.

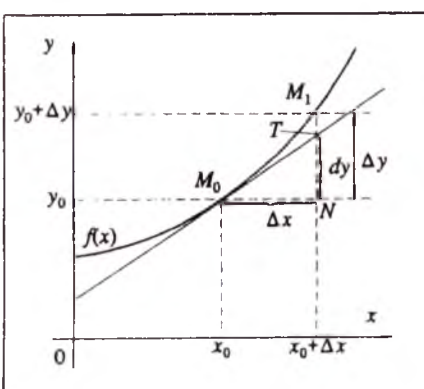


Рис. 3.4а

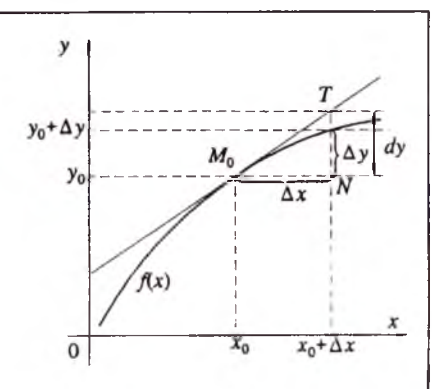


Рис. 3.4б

Приложение дифференциала к приближенным вычислениям. Рассмотрим функцию $y = f(x)$, ее приращение $\Delta f(x_0) = f(x_0+\Delta x) - f(x_0)$ и дифференциал $df(x_0) = f'(x_0)\Delta x^0$ в точке x_0. Выше было установлено, что при достаточно малых Δx^0 имеем

$$\Delta f(x_0) \approx df(x_0).$$

Как правило, вычислять $df(x_0)$ значительно проще, чем $\Delta f(x_0)$ и поэтому на практике последнюю формулу применяют в приближенных вычислениях. К сожалению при этом мы не имеем оценки погрешности.

Пример 3. Найти приближенное значение приращения функции $f(x) = 3x^2 - 7$ при $x_0 = 2$ и $\Delta x^0 = 0{,}001$.

Решение: $\Delta f(x_0) \approx 6x_0\Delta x$ 0,012. Тогда $f(x^0+\Delta x^0) \approx (3(x^0)^2-7)+6x^0\Delta x^0=5+0{,}012=5{,}012$.

3.5. Первообразная и неопределенный интеграл

Напомним, что основная задача дифференциального исчисления заключается в следующем: дана функция $F(x)$, требуется найти ее производную (например, найти предельные издержки, зная суммарные издержки). При этом, если производная существует в каждой точке x некоторого промежутка X, то это также некоторая функция $f(x)$ на X, такая, что $f(x) = F'(x)$. Однако часто приходится решать и обратную задачу: дана функция $f(x)$, требуется найти функцию $F(x)$ такую, что $F'(x) = f(x)$ (например, найти суммарные издержки, зная предельные издержки). Для решения обратной задачи служит операция интегрирования, обратная операции дифференцирования.

Определение. Дифференцируемая функция $F(x)$, определенная на некотором промежутке X, называется первообразной для функции $f(x)$, определенной на том же промежутке, если для всех x из этого промежутка $F'(x) = f(x)$, или, что то же самое $dF(x) = f(x)dx$.

Пример 4. Найти какую-либо первообразную для функции $f(x) = 3x^2$. Функция $F(x) = x^3$ является первообразной для $f(x) = 3x^2$, так как $F'(x) = (x^3)' = 3x^2 = f(x)$. Нетрудно заметить, что первообразная x^3 не является единственной для функции $3x^2$. В самом деле, в качестве первообразной можно было взять и функции: x^3+5, x^3-2 и вообще $x^3 + C$, где C - произвольная постоянная, потому что $(x^3 + C)' = 3x^2$. Приводим формулировку теоремы, выражающей основное свойство первообразных.

Определение. Совокупность всех первообразных для функции $f(x)$, определенных на некотором промежутке X, называется неопределенным интегралом от функции $f(x)$ на этом промежутке и обозначается символом $\int f(x)dx$ (читается: “интеграл от эф от икс де икс"). Если $F(x)$ является первообразной для функции $f(x)$ на промежутке X, то согласно этому определению имеем $\int f(x)dx = F(x) + C$.

Определение. Функция $f(x)$ называется подынтегральной функцией, $f(x)dx$ - подынтегральным выражением, x - переменной интегрирования, символ $\int$ - знаком неопределенного интеграла, C - постоянной интегрирования.

Основные свойства неопределенного интеграла. Пусть функция $F(x)$ является первообразной для функции $f(x)$ на некотором промежутке X, т.е. $F'(x) = f(x)$. Тогда по определению $\int f(x)dx = F(x) + C$. Непосредственно из равенств (1) и (5) следуют свойства:

1) Производная неопределенного интеграла равна подынтегральной функции. Имеем: $(\int f(x)dx)' = (F(x)+C)' = F'(x) = f(x)$.

2) Дифференциал неопределенного интеграла равен подынтегральному выражению. Имеем: $d(\int f(x)dx) = (\int f(x)dx)'dx = f(x)dx$.

3) Неопределенный интеграл от дифференциала некоторой функции равен этой функции плюс произвольная постоянная. Имеем: $\int dF(x) = \int F'(x)dx = f(x)dx = F(x)+C$. Свойства 1) и 2) используют обычно для проверки результатов интегрирования.

4) Постоянный множитель можно выносить за знак интеграла, т.е. если $a = \text{const} \neq 0$, то $\int af(x)dx = a\int f(x)dx$.

5) Неопределенный интеграл от алгебраической суммы двух непрерывных функций равен алгебраической сумме интегралов от этих функций в отдельности, т.е. $\int (f(x) \pm g(x))dx = \int f(x)dx \pm \int g(x)dx$.

ПРИЛОЖЕНИЕ

Вспомогательные сведения

1. Определение предела функции

Пусть функция $y = f(x)$ определена в некоторой окрестности точки x_0, быть может, за исключением самой точки x_0.

Определение. Число A называется пределом функции $f(x)$ при x, стремящемся к x_0 (или в точке x_0), если для любого $\varepsilon > 0$ существует такое $\delta > 0$, что для всех x, удовлетворяющих условиям $|x - x_0| < \delta$, $x \neq x_0$, имеет место неравенство $|f(x) - A| < \varepsilon$.

Если A есть предел функции $f(x)$ при x, стремящемся к x_0, то пишут $\lim\limits_{x \to x_0} f(x) = \mathrm{A}$ или $f(x) \to A$ при $x \to x_0$.

Приведем основные свойства пределов:

1) Постоянный множитель может быть вынесен из-под знака предела.

2) Предел суммы (разности, произведения) равен сумме (разности, произведению) пределов.

2. Определение непрерывности функции

Определение. Функция $f(x)$ называется непрерывной в точке $x = x_0$, если она определена в некоторой окрестности точки x_0 (следовательно, и в самой точке x_0), существует предел функции при $x \to x_0$ и он равен значению функции в этой точке:

$$\lim_{x \to x_0} f(x) = f(x_0).$$

Другими словами, функция $f(x)$ непрерывна в точке $x = x_0$, если для любого $\varepsilon > 0$ существует такое $\delta > 0$, что для всех x, удовлетворяющих условиям $|x - x_0| < \delta$, имеет место неравенство $|f(x) - f(x_0)| < \varepsilon$.

Определение. Функция называется непрерывной на некотором промежутке, если она непрерывна в каждой точке этого промежутка.

Вопросы к главе 3

1. Какие экономические задачи решаются с применением методов дифференциального исчисления?
2. Что представляет собой предельный анализ в экономике?
3. Для решения каких экономических задач недостаточно применения средних величин, а необходимо применение предельных?
4. Каков алгоритм нахождения производной произвольной функции?
5. Что такое дифференциал функции и в каких экономических задачах он используется?
6. Что такое определенный интеграл функции и в каких экономических задачах он используется?

ГЛАВА 4
ПРИМЕНЕНИЕ ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ К ИССЛЕДОВАНИЮ ФУНКЦИЙ

4.1. Возрастание и убывание функций. Признаки возрастания и убывания функции

Рассмотрим вначале самый простой случай - постоянную (на некотором интервале) функцию. Из постоянства функции вытекает равенство нулю ее производной. В этом случае говорят, что равенство нулю производной на некотором интервале есть необходимое условие постоянства функции на этом интервале. Можно легко доказать, что и, наоборот, из равенства нулю производной функции на некотором интервале следует ее постоянство на этом интервале. В этом случае говорят, что постоянство функции на некотором интервале есть достаточное условие равенства нулю производной этой функции на том же интервале.

Перейдем теперь к рассмотрению возрастающих и убывающих функций. Их определения были даны в главе 3. Поэтому мы лишь сформулируем необходимые и достаточные условия возрастания и убывания функций на некотором интервале.

Необходимое условие возрастания функции. Если дифференцируемая функция $y=f(x)$, $x \in (a;b)$, возрастает на интервале $(a;b)$, то $f'(x_0) \geq 0$ для любого $x_0 \in (a;b)$.

Из определения возрастающей функции имеем: для любых $x_0 \in (a;b)$, $x \in (a;b)$ из $x > x_0$ следует, что $f(x) > f(x_0)$, а из $x < x_0$ следует, что $f(x) < f(x_0)$.

В обоих случаях $\dfrac{f(x)-f(x_0)}{x-x_0} > 0$, а следовательно, $\lim\limits_{x \to x_0}\dfrac{f(x)-f(x_0)}{x-x_0} \geq 0$.

т.е. $f'(x_0) \geq 0$.

Необходимое условие убывания функции. Если дифференцируемая функция $y=f(x)$, $x \in (a;b)$, убывает на интервале $(a;b)$, то $f'(x_0) \leq 0$ для любого $x_0 \in (a;b)$.

Доказательство этого утверждения аналогично предыдущему. Достаточные признаки монотонности функции вытекают из следующих двух утверждений, которые мы приводим без доказательства.

Достаточное условие возрастания функции. Если функция, $y=f(x)$, $x \in (a;b)$, имеет положительную производную в каждой точке интервала $(a;b)$, то эта функция возрастает на интервале $(a;b)$.

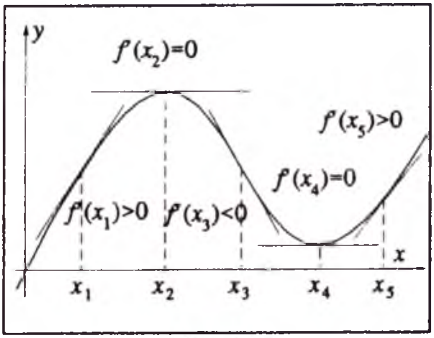


Рис. 4.1а

Достаточное условие убывания функции. Если функция $y=f(x)$, $x \in (a;b)$, имеет отрицательную производную в каждой точке интервала $(a;b)$, то эта функция убывает на интервале $(a;b)$.

Проиллюстрируем эти условия на рис. 4.1а, на котором приведена функция, возрастающая в интервалах $-\infty < x < x_2$ и $x_4 < x < +\infty$ и убывающая в интервале $x_2 < x < x_4$.

4.2. Экстремумы функции. Необходимые и достаточные условия экстремума. Вторая производная и ее геометрическая интерпретация

4.2.1. Понятие экстремума функции

Определение. Точка x_0 из области определения функции $f(x)$ называется точкой минимума этой функции, если найдется такая δ-окрестность $(x_0 - \delta; x_0 + \delta)$ точки x_0, что для всех $x \neq x_0$ из этой окрестности выполняется неравенство $f(x) > f(x_0)$.

Определение. Точка x_0 из области определения функции $f(x)$ называется точкой максимума этой функции, если найдется такая δ-окрестность $(x_0 - \delta; x_0 + \delta)$ точки x_0, что для всех $x \neq x_0$ из этой окрестности выполняется неравенство $f(x) < f(x_0)$.

Определение. Точки минимума и максимума называются точками экстремума, а значения функции в этих точках называются экстремумами функции. Рассмотрим график функции $y=f(x)$, $x \in [a;b]$, (см. рис. 4.1б). Точки x_1 и x_3 являются точками максимума, а x_2 и x_4 точками минимума. Из рис. 4.1б видно, что минимум в точке x_4 больше максимума данной функции в точке x_1. Это объясняется тем, что экстремум функции связан с определенной δ - окрестностью

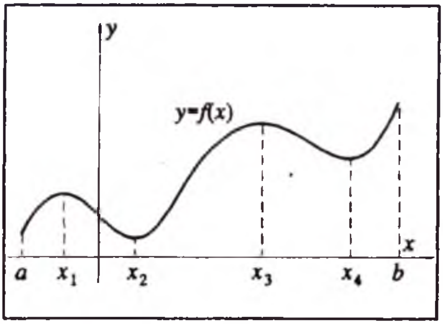


Рис. 4.1б

точки экстремума, а не со всей областью определения функции. По этой причине употребляется термин "локальный экстремум", т.е.экстремум, связанный с данным местом. Этим же объясняется и тот факт,что точки a и b не относятся к точкам экстремума. Для них не существует δ - окрестности, принадлежащей области определения функции.

4.2.2. Необходимые условия существования экстремума

Необходимые условия существования экстремума дает теорема Ферма, которая известна по школьному курсу, поэтому мы приводим лишь ее формулировку.

Теорема Ферма. Если точка x_0 является точкой экстремума функции $y=f(x)$ и в этой точке существует производная $f'(x_0)$, то $f'(x_0)=0$.

Эта теорема имеет простой геометрический смысл: касательная к графику функции $y=f(x)$ в точке, удовлетворяющей условиям теоремы Ферма, параллельна оси абсцисс (см. рис. 4.2)

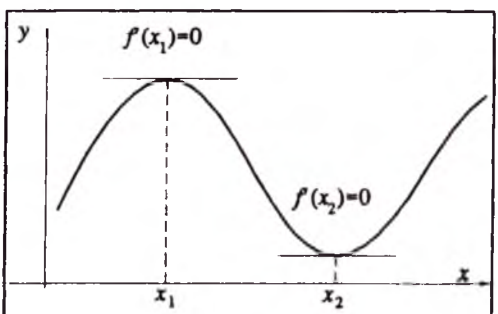


Рис. 4.2

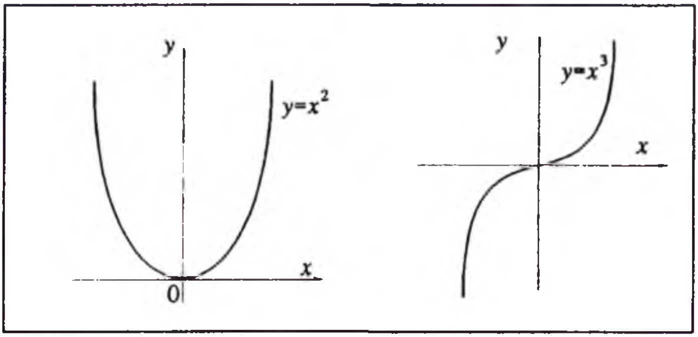


Рис. 4.3а *Рис. 4.3б*

Определение. Точки, в которых производная функции обращается в нуль или не существует, называются критическими точками (первого рода).

Пример 1. Производная функции $f(x) = x^2$ в точке $x_0=0$ обращается в нуль и, как видно из рис. 4.3а, в этой точке данная функция имеет экстремум (минимум). Теорема Ферма дает лишь необходимое условие существования экстремума, но не достаточное.

Пример 2. Производная функции $f(x) = x^3$ в точке $x_0 = 0$ обращается в нуль, а экстремума в этой точке функция не имеет (см. рис. 4.3б) Как показывают следующие примеры, и в тех критических точках, в которых производная не существует, функция также может иметь или не иметь экстремум.

Пример 3. Функция $f(x) = |x|$ в точке $x_0 = 0$ не имеет производной (см. рис. 4.3в). Однако, как видно из рис. 4.3в, в точке $x_0 = 0$ она имеет экстремум (минимум).

Пример 4. Рассмотрим функцию $f(x) = \sqrt[3]{x}$ (см. рис. 4.3г). По графику видно, что в точке $x_0 = 0$ данная функция экстремума не имеет. Производная $f(x) = (\sqrt[3]{x})' = (x^{1/3})' = (x^{-2/3}) = \dfrac{1}{3\sqrt[3]{x^2}}$ в рассматриваемой точке не существует.

Таким образом, экстремум функции, если он существует, может быть только в критических точках. Однако не во всякой критической точке функция имеет экстремум. Чтобы выяснить, в каких критических точках функция имеет экстремум, рассмотрим достаточные условия существования экстремума.

4.2.3. Достаточные условия существования экстремума

Первое достаточное условие экстремума. Пусть функция $y = f(x)$ непрерывна в точке x_0 и в некоторой ее δ-окрестности имеет производную, кроме, быть может, самой точки x_0. Тогда:

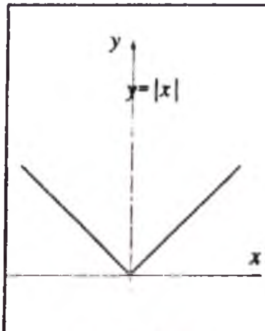


Рис. 4.3б

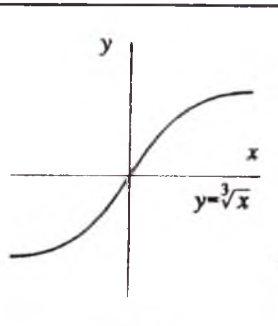


Рис. 4.3в

1) если производная $f'(x)$ при переходе через точку x_0 меняет знак с плюса на минус, то x_0 является точкой максимума.

2) если производная $f'(x)$ при переходе через точку x_0 меняет знак с минуса на плюс, то x_0 является точкой минимума.

3) если производная $f'(x)$ при переходе через точку x_0 не меняет знак, то в точке x_0 функция $f(x)$ не имеет экстремума.

Пример 5. Исследовать на экстремум функцию $y=f(x)=(2x+1)(x-2)^{2/3}$.

1) Находим производную данной функции $f'(x)={}^{10}/_{3}(x-1)(x-2)^{-1/3}$.

2) Находим критические точки:

а) решая уравнение получим $x = 1$.

б) $f(x)$ не существует при $x = 2$.

Следовательно, критические точки: $x_1 = 1$ и $x_2 = 2$.

3) Методом пробных точек определяем знак производной в каждом из интервалов: $(-\infty; 1)$, $(1; 2)$, $(2; +\infty)$ (рис. 4.4) Имеем таким образом $x_1 = 1$ - точка максимума, а $x_2 = 2$ - точка минимума.

4) Вычисляем значения данной функции в точках экстремума $y_{max} = f(1) = 3$, $y_{min} = f(2) = 0$.

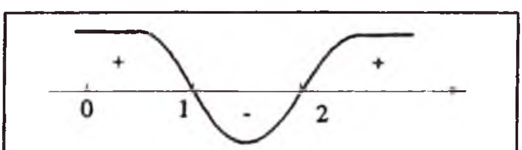


Рис. 4.4

Второе достаточное условие экстремума. Если функция $y = f(x)$ определена и дважды дифференцируема в некоторой окрестности точки x_0, причем $f'(x_0) = 0$, а $f''(x_0)\neq0$, то в точке x_0 функция $f(x)$ имеет максимум, если $f''(x_0) < 0$, и минимум, если $f''(x_0) > 0$.

Пример 6. Исследовать на экстремум функцию $y=(x^3)-2x^2+3x-4$.

1) Находим производную $f'(x)=((x^3)-2x^2+3x-4)'=x^2-4x+3$.

2) Решая уравнение $x^2-4x+3=0$, находим критические точки: $x_1=1$ и $x_2=3$.

3) Находим вторую производную: $f''(x)=(f'(x))'=2x-4$.

4) Определяем знак второй производной в критических точках, для чего вычисляем $f''(1)=-2<0$ и $f''(3)=2>0$. Следовательно, $x_1=1$ - точка максимума, а $x_2=3$ - точка минимума.

5) Вычисляем максимальное и минимальное значения функции $y_{max}=f(1)=-^8/_3$, $y_{min}=f(3)=-4$. Заметим, что в случае, когда вторая производная в критической точке обращается в нуль или не существует, второе правило нахождения экстремума с помощью второй производной неприменимо. В этом случае исследование функции на экстремум можно проводить по первому правилу.

Пример 7. Исследовать на экстремум функцию $f(x)=x^4-2$.

1) Находим $f'(x)=4x^3$.

2) Решая уравнение $4x^3=0$, получаем критическую точку $x=0$.

3) Находим $f''(x)=12x^2$.

4) Вычисляем $f''(0)=0$. В критической точке вторая производная обращается в нуль, поэтому исследование проводим по первому правилу. Так как $f'(x)<0$ при $x<0$, а $f'(x)>0$ при $x>0$, то в точке $x=0$ данная функция имеет минимум, причем $f_{min}=f(0)=-2$.

4.3. Выпуклость графика функции. Точки перегиба

4.3.1. Выпуклость графика функции

При исследовании поведения функции и формы ее графика полезно установить, на каких интервалах график функции обращен выпуклостью вверх, а на каких - выпуклостью вниз. Прежде всего выясним понятие выпуклости графика функции, имеющей на некотором интервале непрерывную производную.

Определение. График функции $y=f(x)$, $x\in(a;b)$ называется выпуклым вверх (вогнутым вниз) на интервале $(a;b)$, если график расположен ниже (точнее не выше) любой своей касательной (см.рис.5). Сама функция $f(x)$ также называется выпуклой вверх (вогнутой вниз).

Определение. График функции $y=f(x)$, $x\in(a;b)$ называется выпуклым вниз (вогнутым вверх) на интервале $(a;b)$, если он расположен выше (точнее не ниже) любой своей касательной (см.рис.4.6). Сама функция $f(x)$ также называется выпуклой вниз (вогнутой вверх).

На интервале выпуклости вверх (вогнутости вниз) производная функции убывает. В самом деле, из рис. 4.5 видно, что с возрастанием аргумента x величина угла α, образованного касательной с положительным направлением оси Ox, убывает, принимая значения

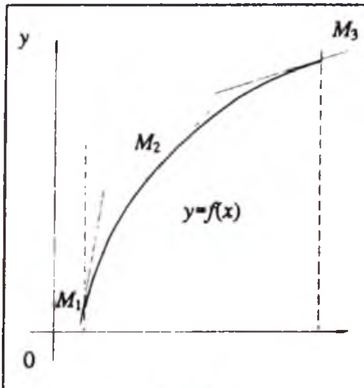


Рис. 4.5

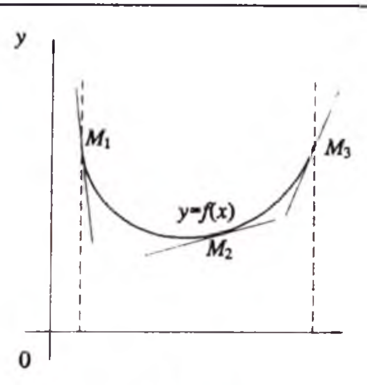


Рис. 4.6

между $\frac{\pi}{2}$ и $-\frac{\pi}{2}$. При этом $\operatorname{tg}\alpha = f'(x)$ также убывает, принимая значения между $+\infty$ и $-\infty$.

Из рис. 4.6 аналогичным образом заключаем, что на интервале выпуклости вниз (вогнутости вверх) производная $f'(x)$ возрастает. Можно показать, что имеют место и обратные утверждения.

Достаточное условие выпуклости графика функции. Если на интервале $(a;b)$ дважды дифференцируемая функция $y = f(x)$, $x \in (a;b)$ имеет отрицательную (положительную) вторую производную, то график функции является выпуклым вверх (вниз).

Допустим для определенности, что $f''(x) < 0$ для всех $x \in (a;b)$. Рассмотрим производную $f'(x)$ как функцию от x, а $f'(x)$ - как ее первую производную. Тогда функция $f'(x)$ убывает на интервале $(a;b)$, а следовательно, по отмеченному выше график функции $y = f(x)$ на этом интервале является выпуклым вверх. Аналогично, если $f''(x) > 0$ для всех $x \in (a;b)$, то график функции $y = f(x)$ на интервале $(a;b)$ является выпуклым вниз.

Исследовать на выпуклость график функции $y = f(x)$ означает найти те интервалы из области ее определения, в которых вторая производная $f''(x)$ сохраняет свой знак. Заметим, что $f''(x)$ может менять свой знак лишь в точках, где $f''(x)=0$ или не существует. Такие точки принято называть критическими точками второго рода.

Пример 8. Исследовать на выпуклость график функции $f(x)= x^3 - 3x^2 + 2x + 1$. Данная функция определена на всей числовой прямой. Находим критические точки второго рода $f'(x) = 3x^2 - 6x + 2$, $f''(x) = 6x - 6$, $6x - 6 = 0$, т.е. $x = 1$. Итак, $x = 1$ - критическая точка второго рода. Методом пробных точек определяем знак $f'(x)$ в каждом из

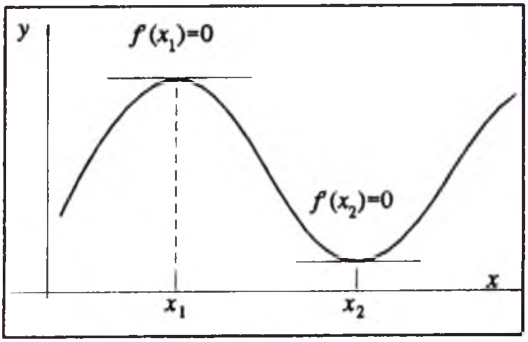


Рис. 4.7

интервалов $(-\infty,1)$ и $(1,+\infty)$. Так, при $x = 0 \in (-\infty,1)$ имеем, $f''(0) = -6 < 0$, а при $x = 2 \in (1,+\infty)$ имеем $f''(2) = 6 > 0$, итак, в точке $x = 1$ производная $f''(x)$ меняет знак с минуса на плюс. Следовательно, на интервале $(-\infty,1)$ график данной функции обращен выпуклостью вверх, а на интервале $(1,+\infty)$ - выпуклостью вниз. (см. рис. 4.7).

4.3.2. Точки перегиба

Определение. Точка графика непрерывной функции $f(x)$, в которой существует касательная и при переходе через которую график функции меняет направление выпуклости, называется точкой перегиба. Согласно определению в точке перегиба касательная к графику функции с одной стороны расположена выше графика, а с другой - ниже, т.е. в точке перегиба касательная пересекает кривую (см. рис. 4.8).

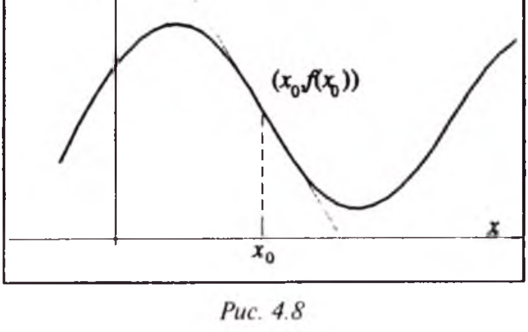


Рис. 4.8

Необходимое условие существования точки перегиба. Если функция $y = f(x)$ имеет непрерывные производные до второго порядка включительно на интервале $(a;b)$ и точка $(x_0, f(x_0))$, где $x_0 \in (a;b)$, является точкой перегиба графика функции $f(x)$, то $f''(x_0) = 0$.

Так как точка $(x_0, f(x_0))$ является точкой перегиба, то слева и справа от x_0 функция $f''(x)$ имеет разные знаки. Но тогда в силу непрерывности второй производной имеем $f''(x_0) = 0$.

Достаточное условие существования точки перегиба. Если функция $y = f(x)$, $x \in (a;b)$ дважды дифференцируема на интервале $(a;b)$ и при переходе через $x_0 \in (a;b)$ вторая производная $f''(x)$ меняет знак, то точка кривой с абсциссой $x = x_0$ является точкой перегиба.

4.4. Асимптоты кривой

Определение. Прямая, имеющая уравнение $y = kx + b$, называется наклонной асимптотой графика функции $y = f(x)$ при $x \to \infty$, если $\lim\limits_{x\to\infty}(f(x) - kx - b) = 0$. Отсюда $k = \lim\limits_{x\to\infty}\frac{f(x)}{x}$, $b = \lim\limits_{x\to\infty}(f(x) - kx)$. Имеет место и обратное: из последних соотношений следует, что прямая $y = kx + b$ является наклонной асимптотой графика функции $y = f(x)$. По выведенным формулам вычисляются угловой коэффициент k и начальная ордината b двух асимптот $y = kx + b$ отдельно при $x \to +\infty$ и при $x \to -\infty$. Очевидно, что если $k = 0$, то уравнение асимптоты примет вид $y = b$.

Определение. Асимптота, определяемая уравнением $y = b$, называется горизонтальной асимптотой.

Определение. Прямая, имеющая уравнение $x = a$, называется вертикальной асимптотой, если $\lim\limits_{x\to a} f(x) = \infty$.

Для определения вертикальных асимптот следует отыскать те значения x, вблизи которых функция $f(x)$ неограниченно возрастает по модулю. Обычно это точки разрыва второго рода данной функции.

Пример 9. Найти асимптоты графика функции $f(x) = \dfrac{x^2+1}{x-2}$.

Так как $\lim\limits_{x\to 2}\dfrac{x^2+1}{x-2} = \pm\infty$, то прямая $x = 2$ является вертикальной асимптотой. Находим

$$k = \lim_{x\to\infty}\frac{f(x)}{x} = \lim_{x\to\infty}\frac{x^2+1}{x(x-2)} = 1,$$

$$b = \lim_{x\to\infty}(f(x) - kx) = \lim_{x\to\infty}\left\{\frac{x^2+1}{x-2} - x\right\} = \lim_{x\to\infty}\frac{2x+1}{x-2} = 2.$$

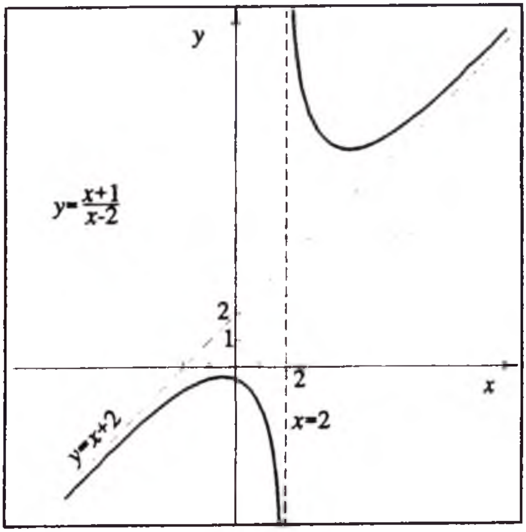


Рис. 4.9

Итак, прямая, имеющая уравнение $y = x + 2$, является наклонной асимптотой графика данной функции при $x \to \pm\infty$. Таким образом, график данной функции имеет вертикальную асимптоту, имеющую уравнение $x = 2$, и наклонную асимптоту, имеющую уравнение $y = x + 2$ (см. рис. 4.9).

Общая схема исследования функций и построения графиков

С учетом изложенного выше можно рекомендовать следующую схему исследования функции и построения ее графика:

1) найти область определения функции;
2) исследовать функцию на четность и нечетность;
3) исследовать функцию на периодичность
4) исследовать функцию на непрерывность, найти точки разрыва;
5) найти критические точки первого рода;
6) найти интервалы монотонности и экстремумы функции;
7) найти критические точки второго рода;
8) найти интервалы выпуклости и точки перегиба;
9) найти асимптоты графика функции;
10) найти точки пересечения графика функции с осями координат (если это возможно);
11) построить график функции.

4.5. Исследование функций в экономике. Нахождение максимума прибыли

В качестве примера рассмотрим задачу выбора оптимального объема производства фирмой, функция прибыли которой может быть смоделирована зависимостью

$$\pi(q) = R(q) - C(q) = q^2 - 8q + 10$$

1. Находим производную этой функции

$$\pi'(q) = R'(q) - C'(q) = 2q - 8$$

2. Приравниваем производную нулю

$$\pi'(q) = 2q - 8 = 0 \to q_{extr} = 4$$

Является ли объем выпуска, равный четырем оптимальным для фирмы? Чтобы ответить на этот вопрос, надо проанализировать характер изменение знака производной при переходе через точку экстремума.

3. Анализируем характер изменения знака производной

При $q < q_{extr} = 4 \to \pi'(q) < 0$ и прибыль убывает.

При $q > q_{extr} = 4 \to \pi'(q) > 0$ и прибыль возрастает.

Следовательно, в точке экстремума $q_{extr} = 4$ прибыль принимает минимальное значение, и таким образом этот объем производства не является оптимальным для фирмы.

4. Принятие решения.

Каким же будет оптимальный объем выпуска для фирмы? Ответ на этот вопрос зависит от дополнительного исследования производственных мощностей фирмы. Если фирма не может производить за рассматриваемый период больше 8 единиц продукции ($p(q=8) = p(q=0) = 10$), то оптимальным решением для фирмы будет вообще ничего не производить, а получать доход от сдачи в аренду помещений и/или оборудования. Если же фирма способна производить за рассматриваемый период больше 8 единиц продукции, то оптимальным решением для фирмы будет выпуск на пределе своих производственных мощностей.

Из этого простого примера видно, насколько важно исследование функций для принятия оптимальных решений, а также для других экономических задач.

Вопросы к главе 4

1. Дайте определение производной.
2. Опишите геометрический смысл производной.
3. Сформулируйте правило вычисления производной суммы двух функций. Пользуясь этим правилом и определением производной, найдите $f'(x)$, если $f(x)=x^2+x$.
4. Сформулируйте правило вычисления производной разности двух функций. Пользуясь этим правилом и определением производной, найдите $f'(x)$, если $f(x)=5x^2-2x$.
5. Сформулируйте правило вычисления производной произведения двух функций. Пользуясь этим правилом и определением производной, найдите $f'(x)$, если $f(x)=(x-2)\cdot(x+3)$.
6. Сформулируйте правило вычисления производной частного двух функций. Пользуясь этим правилом и определением производной, найдите $f'(x)$, если $f(x)=\frac{x-1}{x+4}$.
7. Пользуясь формулами производных элементарных функций и правилами дифференцирования, вычислите производные следующих функций:
$$y=x^4+5x^3;\ y=x^{\frac{1}{4}}-x^{\frac{1}{6}};\ y=\ln(x);\ y=3^x;\ y=e^x.$$
8. Что означает дифференциал функции y от переменной x?
9. Как можно использовать понятие дифференциала для приближенных вычислений? Приведите пример такого использования в экономике.
10. Вычислите, на сколько примерно увеличилась сторона квадрата, если его площадь увеличилась с 4 см^2 до 4,04 см^2.
11. Сформулируйте достаточное условие возрастания (убывания) функции на отрезке.
12. Определите промежутки строгого возрастания (убывания) следующих функций:
$$y=2+x-x^2;\ \ y=3x-x^3;\ \ y=e^x;\ \ y=\ln(x).$$
13. Дайте определение экстремума функции. Приведите пример использования понятия экстремума в экономике.
14. Сформулируйте достаточное условие экстремума.
15. Сформулируйте необходимые условия экстремума.
16. Исследуйте на экстремум следующие функции:
$$y=2+x+x^2;\ y=(x-3)^3;\ y=x+\frac{1}{x};\ y=\frac{2x}{1+x^2}.$$
17. Что такое вторая производная?
18. Сформулируйте достаточные условия выпуклости (вогнутости) функции на отрезке. Приведите примеры экономических зависимостей, описываемых выпуклыми (вогнутыми) функциями.

19. Сформулируйте достаточное условие точки перегиба. Приведите пример экономической зависимости, описываемой функцией, имеющей точку перегиба.
20. Найдите промежутки выпуклости (вогнутости) и точки перегиба графиков следующих функций:

$$y=3x^2-x^3;\ y=x+x^{\frac{5}{3}}\ y=\ln(1+x^2).$$

21. Постройте графики зависимости издержек и дохода от объема производства. Укажите на них значения объемов производства, при которых:

   а) прибыль максимальна,

   б) убытки максимальны.

ГЛАВА 5
ЭЛАСТИЧНОСТЬ И ЕЕ ПРИМЕНЕНИЕ В ЭКОНОМИЧЕСКОМ АНАЛИЗЕ

Важнейшим направлением применения дифференциального исчисления в экономике является введение с его помощью понятия эластичности. Коэффициент эластичности показывает относительное изменение исследуемого экономического показателя под действием единичного относительного изменения экономического фактора, от которого он зависит при неизменных остальных влияющих на него факторах.

5.1. Эластичность функции и ее геометрический смысл

Пусть величина y зависит от x, и эта зависимость описывается функцией $y=f(x)$. Изменение независимой переменной x (Δx) приводит в силу функциональной зависимости к изменению переменной y (Δy). Встает вопрос, как измерить чувствительность зависимой переменной y к изменению x. Одним из показателей реагирования одной переменной на изменение другой служит производная

$$y_x' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

характеризующая скорость изменения функции с изменением аргумента x. Однако в экономике этот показатель неудобен тем, что он зависит от выбора единиц измерения.

Например, если мы рассмотрим функцию спроса на сахар (Q) от его цены (P), то увидим, что значение производной при каждой цене P (измеряемой в рублях)

$$Q_p' = \lim_{\Delta p \to 0} \frac{\Delta Q}{\Delta P}$$

зависит от того, измеряется ли спрос на сахар в килограммах или в центнерах. В первом случае производная измеряется в кг/руб., во втором - в ц/руб., соответственно ее значение при одном и том же значении цены будет различным в зависимости от единиц измерения величины спроса. Поэтому для измерения чувствительности изменения функции к изменению аргумента в экономике изучают связь не абсолютных изменений переменных x и y (Δx и Δy), а их относительных или процентных изменений.

Эластичностью функции $y=f(x)$ называется предел отношения относительных изменений переменных y и x.

Если эластичность изменения переменной y при изменении переменной x обозначить $E_x(y)$, то, используя определение производной, получаем

$$E_x(y) = \lim_{\Delta x \to 0} \left[\frac{\Delta y}{y}\right] \Big/ \left[\frac{\Delta x}{x}\right] = \lim_{\Delta x \to 0} \left[\frac{\Delta y}{\Delta x} \cdot \frac{x}{y}\right] = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \cdot \frac{x}{y},$$

$$E_x(y) = \frac{dy}{dx} \cdot \frac{x}{y} = f'(x) \cdot \frac{x}{y} = \frac{f'(x)}{\frac{y}{x}} = \frac{f'(x)}{\frac{f(x)}{x}} = \frac{Mf}{Af},$$

где Mf - маржинальное значение функции f в точке x (см. ниже гл. 6), Af - среднее значение функции в точке x. Эту эластичность называют также предельной или точечной эластичностью.

Т.е. эластичность может быть выражена в виде отношения предельной (Mf) и средней (Af) величин.

Так как $d\ln y = \frac{dy}{y}$, а $d\ln x = \frac{dx}{x}$, то эластичность можно представить в форме "логарифмической производной" $E_x(y) = \frac{d\ln y}{d\ln x}$.

Геометрическая интерпретация эластичности.

Подобно производной, эластичность имеет простую геометрическую интерпретацию.

Рассмотрим убывающую вогнутую функцию $y = f(x)$ (рис. 5.1)

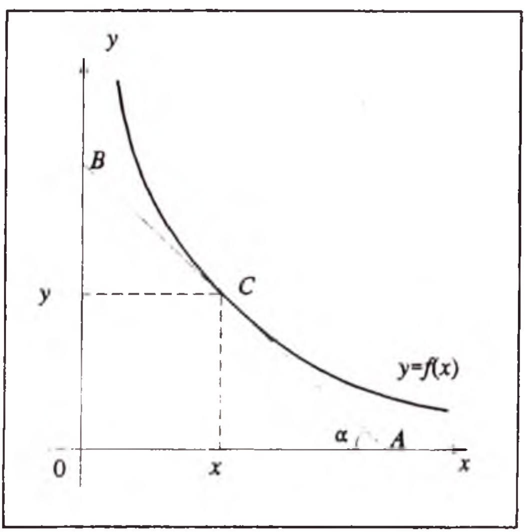


Рис. 5.1

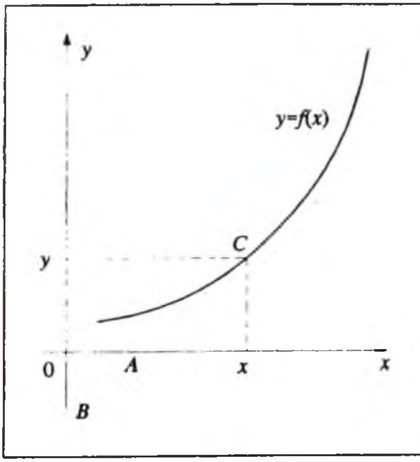


Рис. 5.2

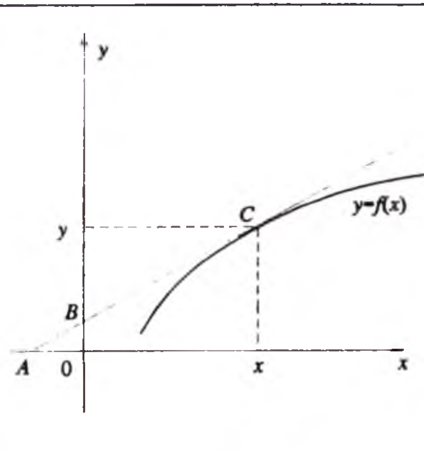


Рис. 5.3

Найдем эластичность этой функции в произвольной точке C с координатами (x,y). Для этого проведем касательную AB к функции $y = f(x)$ в точке C.

Из $\Delta ACX \quad AX = \dfrac{CX}{\operatorname{tg}\alpha}$.

Т.к. производная функции $y = f(x)$ в точке C равна $\operatorname{tg}(180 - \alpha)$, то $\operatorname{tg}\alpha = -f'(x)$.

Следовательно, $AX = \dfrac{f(x)}{-f'(x)} = -\dfrac{f(x)}{f'(x)}$.

Из подобия треугольников CBY и CAX следует, что

$$\frac{CB}{CA} = \frac{CY}{AX} = \frac{OX}{AX} = -\frac{f'(x)x}{f(x)} = -E_x(y).$$

Таким образом, $E_x(y) = -\dfrac{CB}{CA}$,

т.е. геометрически эластичность убывающей функции равна отношению расстояний по касательной от точки C с координатами $(x, f(x))$ до ее пересечения с осями Y и X, взятому, соответственно, со знаком "-".

В случае выпуклой и вогнутой возрастающих функций (рис. 5.2 и рис. 5.3) эластичность по абсолютной величине также будет равна отношению $\dfrac{CB}{CA}$, а знак эластичности будет определяться направлением отрезков CB и CA. Если точки A и B лежат по одну сторону от точки C на касательной, как на рисунках 5.2, 5.3, то в формуле надо выбрать знак "+". Если A и B лежат по разные стороны, от т. C, как на первом рисунке, то в формуле надо выбрать знак "-" (доказательство для двух последних рисунков вы можете провести самостоятельно).

Отметим также, что эластичность функции, изображенной на рис. 5.2, больше единицы (так как $CB > CA$), а на рис. 5.3 - меньше единицы (так как $CB < CA$).

Дискретный случай.

В дискретном случае, а также при приближенном определении эластичности по дискретному набору данных, определение эластичности уже не столь однозначно, как в непрерывном случае, поскольку в относительном изменении $\delta x = \frac{\Delta x}{x} = \frac{x_2 - x_1}{x}$ не ясно, что брать в качестве x: первоначальное значение ($x = x_1$), конечное значение ($x = x_2$) или среднее значение $x = \frac{x_1 + x_2}{2}$.

В зависимости от этого выбора различают:

конечную (процентную) эластичность

$$E_x(y) = \left(\frac{y_2 - y_1}{y_1}\right) \Big/ \left(\frac{x_2 - x_1}{x_1}\right),$$

среднюю (дуговую) эластичность

$$E_x(y) = \left(\frac{2(y_2 - y_1)}{y_1 + y_2}\right) \Big/ \left(\frac{2(x_2 - x_1)}{x_1 + x_2}\right),$$

а также логарифмическую эластичность

$$E_x(y) = \frac{\Delta \ln y}{\Delta \ln x} = \frac{\ln y_2 - \ln y_1}{\ln x_2 - \ln x_1} = \ln\left(\frac{y_2}{y_1}\right) \Big/ \ln\left(\frac{x_2}{x_1}\right).$$

Все эти выражения мало отличаются друг от друга при небольших относительных (процентных) изменениях величин x и y.

Отметим, что для всех эластичностей используется один и тот же символ $E_x(y)$, ибо из контекста бывает ясно, о какой эластичности идёт речь.

5.2. Свойства эластичности и эластичность элементарных функций

Свойства эластичности:

1. Эластичность - безразмерная величина, значение которой не зависит от того, в каких единицах измерены величины y и x: $E_{ax}(by) = E_x(y)$.

$$E_{ax}(by) = \frac{d(by)}{d(ax)} \cdot \frac{ax}{by} = \frac{b(dy)}{a(dx)} \cdot \frac{ax}{by} = \frac{dy}{dx} \cdot \frac{x}{y} = E_x(y).$$

2. Эластичности взаимно обратных функций - взаимно обратные величины:

$$E_x(y) = \frac{1}{E_y(x)} \Leftarrow E_x(y) = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{1}{\frac{dx}{dy} \cdot \frac{y}{x}} = \frac{1}{E_y(x)}.$$

Например, эластичность величины спроса по цене обратна эластичности цены по величине спроса $\left(E_p(Q) = \frac{1}{E_Q(p)} \right)$.

3. Эластичность произведения двух функций $u(x)$ и $v(x)$, зависящих от одного и того же аргумента x, равна сумме эластичностей: $E_x(uv) = E_x(u) + E_x(v)$.

$$E_x(uv) = \frac{d(uv)}{dx} \cdot \frac{x}{uv} = \frac{v\left(\frac{du}{dx}\right) + u\left(\frac{dv}{dx}\right)}{uv} \cdot x = \frac{du}{dx} \cdot \frac{x}{u} + \frac{dv}{dx} \cdot \frac{x}{v} =$$
$E_x(u)+E_x(v)$.

4. Эластичность частного двух функций $u(x)$ и $v(x)$, зависящих от одного и того же аргумента x, равна разности эластичностей

$$E_x\left(\frac{u}{v}\right) = \frac{d\frac{u}{v}}{dx} \cdot \frac{x}{\frac{u}{v}} = \frac{vdu - udv}{v^2} \cdot \frac{xv}{u} = \frac{du}{dx} \cdot \frac{x}{u} - \frac{dv}{dx} \cdot \frac{x}{v} = E_x(u) - E_x(v).$$

5. Эластичность суммы двух функций $u(x)$ и $v(x)$ может быть найдена по формуле:

$$E_x(u+v) = \frac{d(u+v)}{dx} \cdot \frac{x}{u+v} = \left(\frac{du}{dx} + \frac{dv}{dx}\right) \cdot \frac{x}{u+v} = \frac{uE_x(u) + vE_x(v)}{u+v}.$$

Эластичности элементарных функций:

1. Эластичность степенной функции $y = x^\alpha$ постоянна и равна показателю степени α: $E_x(x^\alpha) = \alpha$.

$$E_x(x^\alpha) = \frac{dx^\alpha}{dx} \cdot \frac{x}{x^\alpha} = \frac{\alpha x^{\alpha-1} \cdot x}{x^\alpha} = \alpha.$$

2. Эластичность показательной функции $y = a^x$ пропорциональна x: $E_x(a^x) = x \ln(a)$.

$$E_x(x^\alpha) = \frac{da^x}{dx} \cdot \frac{x}{a^x} = a^x \cdot x \frac{\ln a}{a^x} = x \cdot \ln a.$$

3. Эластичность линейной функции $y = ax+b$ $E_x(ax+b) = \frac{ax}{ax+b}$.

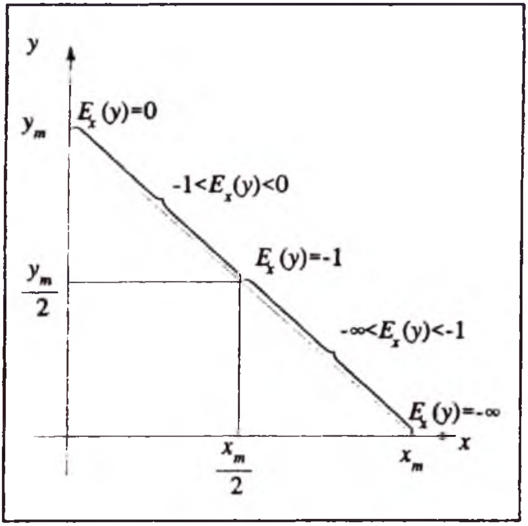


Рис. 5.4

$$E_x(ax+b) = \frac{d(ax+b)}{dx} \cdot \frac{x}{ax+b} = \frac{ax}{ax+b}.$$

Если график линейной функции имеет отрицательный наклон ($a<0$), то эластичность функции меняется от нуля в точке y_m пересечения графиком оси y до минус бесконечности ($-\infty$) в точке пересечения оси x, проходя через значение (-1) в средней точке. Таким образом, хотя прямая имеет постоянный наклон, ее эластичность зависит не только от наклона, но и от того, в какой точке x мы ее находим (рис. 5.4). Функция с бесконечной эластичностью во всех точках называется совершенно эластичной, с нулевой эластичностью во всех точках - совершенно неэластичной.

5.3. Применение эластичности в экономическом анализе

5.3.1. Виды эластичностей в экономике

- Эластичность спроса по цене (прямая)

$$E_p(q) = \left(\frac{dq}{q}\right) \Big/ \left(\frac{dp}{p}\right) = \frac{dq}{dp} \cdot \frac{p}{q},$$

показывающая относительное изменение (выраженное в процентах) величины спроса на какое-либо благо при изменении цены этого блага на один процент и характеризующая чувствительность потребителей к изменению цен на продукцию. Если ценовая элас-

$E_p(q)$	Эластичный спрос		Неэластичный спрос	
$-\infty$		-1		0

тичность спроса по абсолютной величине больше единицы, то спрос называют эластичным (совершенно эластичным при бесконечно большой величине эластичности спроса). Если ценовая эластичность спроса по абсолютной величине меньше единицы, то спрос называют неэластичным (совершенно неэластичным при нулевой эластичности спроса).

И, наконец, если ценовая эластичность спроса по абсолютной величине равна единице, то говорят о спросе с единичной эластичностью.

- *Эластичность спроса по доходу*

$$E_I(q) = \left[\frac{dq}{q}\right] \Big/ \left[\frac{dI}{I}\right] = \frac{dq}{dI}\cdot\frac{I}{q},$$

характеризующая относительное изменение (в процентах) величины спроса на какое-либо благо при изменении дохода потребителей этого блага на один процент. Положительная эластичность спроса по доходу характеризует нормальные (качественные) товары, а отрицательная величина - малоценные (некачественные) товары.

Так, высокий положительный коэффициент спроса по доходу в отрасли указывает, что ее вклад в экономический рост больше, чем доля в структуре экономики, и она имеет шансы на расширение и процветание в будущем. Наоборот, если коэффициент эластичности спроса на продукцию отрасли по доходу имеет небольшое положительное или отрицательное значение, то ее может ожидать застой и перспектива сокращения производства.

- *Перекрестная эластичность спроса по цене*

$$E_{pj}(q_i) = \left[\frac{dq_i}{q_i}\right] \Big/ \left[\frac{dp_j}{p_j}\right] = \frac{dq_i}{dp_j}\cdot\frac{p_i}{q_j},$$

характеризующая относительное изменение (в процентах) величины спроса на одно благо при изменении цены на другое благо (замещающее или дополняющее его в потреблении) на один процент. Положительный знак перекрестной эластичности спроса по цене свидетельствует о замещаемости благ, а отрицательный - о дополняемости.

- *Ценовая эластичность ресурсов*

$$E_{pi}(R_i) = \left[\frac{dR_i}{R_i}\right] \Big/ \left[\frac{dp_i}{p_i}\right] = \frac{dR_i}{dp_i} \cdot \frac{p_i}{R_i},$$

характеризующая относительное изменение (в процентах) величины спроса на какой-либо ресурс (например, труд) при изменении цены этого ресурса (соответственно, заработной платы) на один процент.

- *Эластичность замещения одного ресурса другим*

$$E_{Rj}(R_i) = \left[\frac{dR_i}{R_i}\right] \Big/ \left[\frac{dR_j}{R_j}\right] = \frac{dR_i}{dR_j} \cdot \frac{R_j}{R_i},$$

характеризующая необходимое изменение (в процентах) величины одного ресурса (например, капитала) при изменении количества другого ресурса (например, труда) на один процент с тем, чтобы выпуск при этом не изменился.

5.3.2. Факторы, определяющие эластичность спроса

1. Замещаемость блага в потреблении

Эластичность спроса по цене тем выше, чем выше замещаемость блага.

Замещаемость блага обычно характеризуется наличием и количеством заместителей, а также степенью агрегированности блага. Чем больше у потребителей возможностей заместить потребление данного блага потреблением других благ, тем выше эластичность спроса на это благо. Степень агрегированности блага определяется широтой определения данного блага, тем какое количество разнородных благ входит в понятие данного блага. Например, благо "молочные продукты" включает в себя молоко, кефир, ряженку, простоквашу и другие продукты. Чем выше степень агрегированности блага, тем меньше у него субститутов (и тем меньше у потребителей возможностей заместить потребление данного блага потреблением других благ), и тем ниже эластичность спроса на это благо. Например, эластичность спроса на моющие средства ниже, чем на стиральный порошок, а эластичность спроса на мыло вообще ниже, чем эластичность спроса на мыло конкретной марки.

Правильное определение степени агрегированности блага (с целью определения его эластичности) особенно важно при определении ценовой и налоговой политики. Так, например, эластичность водки относительно низкая и казалось бы, что увеличение акцизов на нее должно привести к увеличению поступлений в бюджет (поскольку при неэластичном спросе выручка растет с увеличением цены). Однако происшедшее в декабре 1993 года повышение ставки акцизов до 90 процентов привело к существенному снижению спроса на отечественные ликеро-водочные изделия, потере конкурентоспособности отрасли по цене и резкому сокращению доходов в бюджет. Причина - неправильное определение степени агрегированности водки, которая должна включать в себя не только отечественную водку, но и импортную (в том числе и из стран ближнего зарубежья). Именно продажа импортной водки и заняла в это время преимущественное место в торговой сети. Спустя несколько месяцев председателем правительства РФ Виктором Черномырдиным было подписано новое постановление, которым снижалась ставка акцизов на отечественную водку до 85% и, одновременно, повышалась ставка акцизов до 250% на импортную водку.

Подобные провалы правительственной политики происходили не только в России, но и в странах с развитой рыночной экономикой (и экономической теорией). Так, например, введение в 80-е годы 6% налога на бензин в Вашингтоне (округ Колумбия), эластичность спроса на который по оценкам экономистов составляла 0.2, привело к 33% падению спроса (что соответствует эластичности 5.5) и через 2 месяца налог был отменен. Причина этого - "узкое" определение бензина в штате Washington D.C., не включившее в себя бензин из соседних штатов Meriland и Virginia, которым потребители и стали заменять подорожавший в Вашингтоне бензин.

2. Удельный вес в доходе

Эластичность спроса по цене тем выше, чем выше удельный вес расходов на данное благо в доходе потребителя. Например, спрос потребителя на спички, практически не изменится, даже если их цена возрастет в несколько раз, что свидетельствует о его низкой эластичности.

3. Субъективная необходимость

Эластичность спроса по цене тем выше, чем ниже субъективная необходимость в данном благе. Обычно считают, что спрос на предметы роскоши более эластичен, чем спрос на предметы первой необходимости. Это не совсем правильно, поскольку решающим фактором здесь является именно субъективная необходимость в данном благе, которая на отдельные предметы роскоши может в силу моды, традиций или других причин может быть достаточно высокой и приводить к низкой эластичности спроса на него. Примером этому служит спрос на цветы 8 марта или 1 сентября.

4. Фактор времени

Эластичность спроса по цене обычно выше, чем больше промежуток времени. Другими словами, долгосрочная эластичность спроса предполагается выше, чем краткосрочная эластичность. Это обычно обосновывается тем, что за долгосрочный промежуток времени потребители могут изменить привычки и найти больше заменителей данному благу.

Однако, при этом не учитывается формирование запаса и время износа блага, оказывающие существенное влияние на решения потребителей и действующие иногда в сторону понижения эластичности с течением времени, особенно для товаров длительного пользования, а также товаров первой необходимости в периоды резкого повышения цен. Например, запасы круп, макаронных изделий, консервов и других товаров, сделанные домашними хозяйствами в России в декабре 1991 года до резкого повышения цен, привели к резкому сокращению спроса на эти товары в начале следующего года и, следовательно, большой краткосрочной эластичности спроса. С течением времени запасы стали истощаться и эластичность спроса на эти товары уменьшилась.

5.3.3. Связь эластичности с выручкой продавцов (расходами покупателей)

Эластичность выручки от продаж какого-либо блага тесно связана с эластичностью спроса на это благо. Используя формулу для выручки $R=pq$ и формулу для эластичности произведения функций, получаем

$$E_p(R) = E_p(q) + E_p(p) = E_p(q) + 1 = 1 - |E_p(q)|,$$

так как эластичность спроса по цене всегда отрицательна (поскольку $p'(q)<0$).

Из полученной формулы видно, что эластичность выручки по цене отрицательна ($E_p(R)<0$) для товаров, спрос на которые эластичен ($|E_p(q)|>1$), и положительна ($E_p(R)>0$) для товаров, спрос на которые неэластичен ($|E_p(q)|<1$). Это означает, что если спрос неэластичен, то изменение цены вызывает изменение выручки в том же направлении и продавцам выгодно повышать цену (что приводит к увеличению их выручки). Для эластичного спроса изменение выручки происходит в направлении, противоположном изменению цены и для повышения выручки продавцам выгодно понижать цену. Аналогично, повышение налога на товар с эластичным спросом повлечет за собой сокращение дохода от налогообложения.

При эластичном спросе выручка растет с увеличением количества или уменьшением цены, а при неэластичном - падает. Например, доходы фермеров сократятся при хорошем урожае, поскольку эластичность спроса на сельскохозяйственную продукцию достаточно низка. Аналогично, повышение цен на государственных предприятиях с целью увеличения поступлений в бюджет, например, повышение цен на железнодорожные билеты, может привести к сокращению поступлений в бюджет, если спрос на соответствующий товар или услугу окажется эластичным.

5.3.4. Связь цены и предельных издержек монополиста

Мы знаем, что совершенно конкурентная фирма (т.е. фирма, функционирующая в условиях совершенной (чистой) конкуренции, устанавливает цену на свою продукцию, равную предельным[1] издержкам: $p_c = MC$. Монополист же назначает цену на свою продукцию выше предельных издержек: $p_m = MC(1 + s)$, где s - надбавка к издержкам, которая может составлять 10%, 20%, 30%,.... Возникает вопрос, как выбрать величину этой надбавки. При более детальном рассмотрении этого вопроса оказывается, что величина надбавки тесно связана с эластичностью.

Действительно, записывая условие максимизации прибыли, как разницы между выручкой и издержками ($\pi = R - C$):

$$\pi' = R' - C' = MR - MC = 0$$

вычислим предельную выручку

$$MR \equiv R'(q) = (p(q) \cdot q)' = p(q) + qp'(q) =$$
$$= p \cdot \left[1 + \frac{q \cdot p'(q)}{p}\right] = p \cdot \left[1 + \frac{1}{E^D}\right] = p \cdot \left[1 - \frac{1}{|E^D|}\right]$$

и приравняем ее к предельным издержкам ($MR = MC$). Получим соотношение

$$p = \frac{MC}{1 - \frac{1}{|E^D|}},$$

из которого следует, что надбавка к предельным издержкам в цене должна быть тем меньше, чем выше эластичность спроса.

Эта же формула позволяет объяснить, как происходит сегментация рынка монопольным производителем с целью дискриминации потребителей и получения от этого дополнительной прибыли. Обычно мы рассматриваем однородный рынок какой-либо продукции на котором все покупатели платят за единицу товара одну и ту же цену. Однако, если монополист может устойчиво разделить поку-

---

[1] Определение предельной величины см. ниже в главе 6, п.6.2.

пателей по какому-либо признаку на две или большее число групп, например, выделяя в отдельную группу студентов при покупке железнодорожных или авиабилетов, то ему выгоднее установить для различных групп различные цены и, таким образом, сегментировать рынок. При этом суммарная выручка от продаж на двух рынках одного товара (или услуги) будет максимальна при равенстве предельных доходов от каждого из рынков (в противном случае было бы выгодно перераспределить объем продаж в пользу рынка с большим предельным доходом). Таким образом,

$$MR_1 = p_1 \cdot \left(1 - \frac{1}{|E_1^D|}\right) = p_2 \cdot \left(1 - \frac{1}{|E_2^D|}\right) = MR_2.$$

Отсюда

$$\frac{p_1}{p_2} = \frac{1 - \frac{1}{|E_2^D|}}{1 - \frac{1}{|E_1^D|}}$$

Следовательно, те покупатели, спрос которых на товар менее эластичен будут платить за него большую цену.

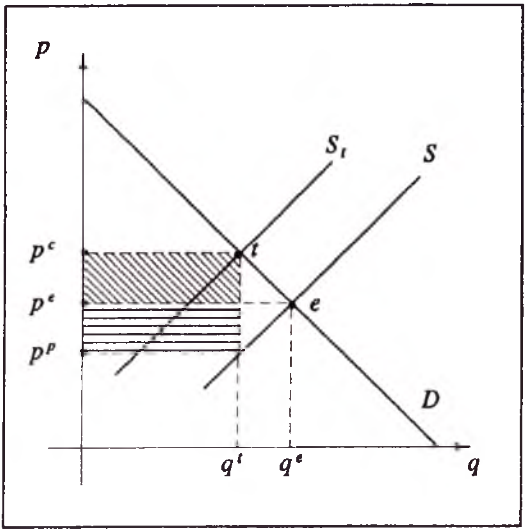
5.3.5. Эластичность и налоговая политика

Когда правительство вводит те или иные налоги на какие-либо товары оно должно иметь ответы на следующие вопросы: На какие товары вводить налог?

С кого взимать налог - с производителей или потребителей? Какова будет величина дополнительных поступлений в бюджет? На кого ляжет основное налоговое бремя? Если налог уже взимается, то стоит ли увеличивать налоговую ставку для покрытия дефицита бюджета?

Интуитивно казалось бы, что основное налоговое бремя ляжет на тех, с кого будут взимать налог и что чем больше будет налоговая ставка, тем больше будут поступления от налогов в бюджет. Более детальный экономический анализ показывает, что величина налогового бремени определяется не формальными плательщиками налога, а величинами эластичности спроса и предложения. Аналогично, увеличение налоговой ставки, эквивалентное увеличению цены, облагаемого налогом товара может привести как к увеличению налоговых поступлений в бюджет, так и к их уменьшению, опять же в зависимости от эластичности.

Для того чтобы разобраться в этих вопросах, рассмотрим более детальную модель взимания налога, основанную на концепции спроса и предложения. Предположим, вначале что налог взимается с производителей, и для простоты будем считать, что налог с единицы продукции t постоянен и не зависит от величины выпуска (это не так, если налог определяется в процентах с выпуска или объема продаж). В этом случае введение налога приводит к параллельному



сдвигу кривой предложения на величину налоговой ставки t.

Из этого рисунка видно, что при введении налога рыночная цена товара повышается от p^e до p^c, которая теперь отличается от цены производителей p^p на величину налога t, а объем продаж уменьшается от q^e до q^t. Суммарная величина налоговых поступлений в бюджет T определяется как произведение налоговой ставки t на объем продаж q^t:

$$T = t \cdot q^t.$$

Одновременно это же выражение определяет и величину налогового бремени, часть которого

$$T_c = q_t \cdot (p^c - p^e)$$

падает на плечи потребителей, а другая часть

$$T_p = q_t \cdot (p^e - p^p)$$

- на производителей.

Нетрудно показать, что сумма этих частей равна налоговым поступлениям в бюджет:

$$T_c + T_p = T = q_t \cdot (p^c - p^p),$$

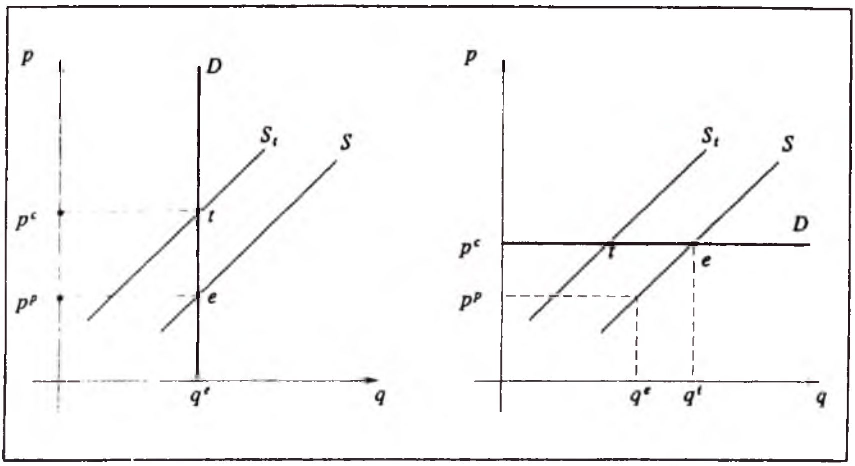
а соотношение этих частей обратно пропорционально соотношению эластичностей спроса и предложения.

Это соотношение следует из определений эластичностей спроса и предложения, отношение которых и дает искомое выражение

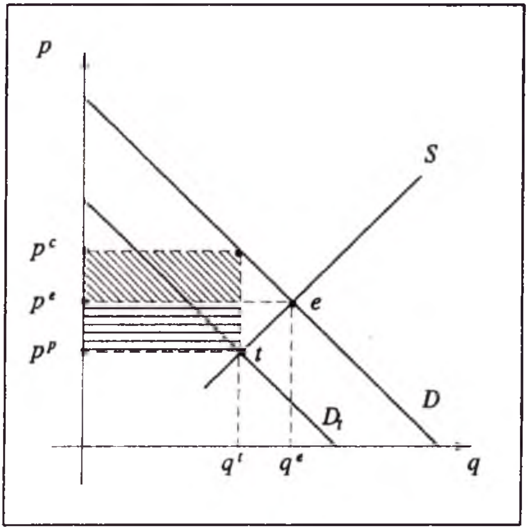
$$\frac{T_c}{T_p} = \frac{p^c - p^e}{p^e - p^p} = -\frac{E^S}{E^D}$$

$$-E^D = \left|-\frac{q_t - q_e}{q_e}\right| \Big/ \left|\frac{p_c - p_e}{p_e}\right|; \quad E^S = \left|\frac{q_t - q_e}{q_e}\right| \Big/ \left|\frac{p_p - p_e}{p_e}\right|.$$

Анализируя его, мы видим что бульшее налоговое бремя падает на экономического агента с меньшей эластичностью, у которого меньше возможностей для ухода от налогового бремени. В частности, если эластичность спроса равна нулю, то все налоговое бремя ляжет на плечи потребителей, так как независимо от величины налога (а следовательно, и от величины цены) потребители не изменят объема покупок. Если же спрос на какой-либо товар характеризуется совершенной эластичностью, то в проигрыше оказываются производители, так как потребители уходят от налога, снижая величину спроса и переходя к потреблению товаров-субститутов. В этом случае все налоговое бремя падает на плечи производителей:



Аналогично происходит и перераспределение налогового бремени в случае, когда налог формально взимается с потребителей. Например, оплачивая какую-либо покупку, покупатель платит по дополнительному чеку определенную сумму или процент от суммы покупки государству. В этом случае введение налога приводит к сдвигу кривой спроса влево:



Сравнивая этот рисунок с рисунком, описывающим ситуацию взимания налога с производителей, можно заметить, что распределение налогового бремени между потребителями и производителями происходит также, как и в предыдущем случае, и опять обратно пропорционально их эластичностям. Таким образом, формальные и фактические плательщики налога не совпадают. Независимо от того, кто является формальным плательщиком налога, фактическим плательщиком оказывается экономический агент с меньшей эластичностью, особенно если эластичности спроса и предложения сильно различаются.

Рассматривая вопрос о влиянии величины налоговой ставки на величину налоговой выручки, нетрудно заметить, что эти величины связаны между собой примерно так же, как связаны выручка от продаж и цена товара. Рассуждая аналогично выводу связи выручки и эластичности, можно получить формулу

$$E_t(T) = \frac{t}{T}\cdot\frac{dT}{dt} = 1 - \frac{\frac{t}{p^e}}{\frac{1}{|E^D|} + \frac{1}{E^S}}$$

Из этой формулы видно, что налоговая выручка возрастает с увеличением налоговой ставки только до тех пор, пока доля ставки налога в цене товара меньше суммы обратных эластичностей спроса и предложения. Это дает возможность устанавливать высокие ставки налогообложения (существенно превышающие цену товара) на товары, спрос на которые неэластичен (или предложение которых неэластично). Примером этому служат акцизы на винно-водочные и табачные изделия.

Таким образом, эластичность спроса важна при принятии ценовых решений производителями, бизнесменами, владельцами стадионов, кинотеатров и других заведений, разработчиками государственной политики и другими экономическими субъектами.

Вопросы к главе 5

1. Что показывает в экономике коэффициент эластичности?
2. Что такое эластичность функции?
3. Объясните геометрический смысл эластичности убывающей вогнутой функции.
4. Что такое точечная эластичность, дуговая эластичность? В каких случаях используется каждое из этих понятий?
5. Перечислите свойства эластичности.
6. Как по коэффициенту перекрестной эластичности спроса на два товара определить, являются ли эти товары взаимозаменяемыми или взаимодополняющими?
7. Как с помощью коэффициента эластичности спроса на товар по доходу определить, ожидает ли выпускающую его отрасль процветание или застой?
8. Перечислите экономические приложения понятия эластичности.

ГЛАВА 6
СООТНОШЕНИЯ МЕЖДУ СУММАРНЫМИ, СРЕДНИМИ И ПРЕДЕЛЬНЫМИ ВЕЛИЧИНАМИ В ЭКОНОМИКЕ

6.1. Абсолютные и относительные величины в экономическом анализе

Все экономические показатели можно, с известной долей условности, разделить на абсолютные и относительные. Первые выражаются в каких-либо объемных или денежных единицах и могут быть либо потоковыми (то есть величина за определенный период), либо запасовыми (то есть величина на определенную дату). Вторые, относительные, показатели представляют собой отношения абсолютных (или других относительных) показателей, то есть количество единиц одного показателя на одну единицу другого. Относительными показателями являются не только соотношения разных показателей в один и тот же момент времени, но и одного и того же - в разные моменты; это - темпы роста данного показателя. В экономическом анализе и принятии решений в одних случаях важны абсолютные показатели (например, общий объем прибыли), в других - относительные (например, доход на душу населения).

Как правило, для комплексного анализа экономической ситуации, для выбора наилучшего решения важны как абсолютные, так и относительные показатели. Пусть, например, фирма решает вопрос о необходимом масштабе расширения (или сокращения) объема производства. Ее, естественно, интересует (абсолютный) показатель прибыли, являющийся разностью двух других таких показателей - выручки и издержек. Но при решении поставленной задачи максимизации прибыли фирма широко использует два типа относительных показателей: это *средние* и *предельные* величины (прибыли, выручки, издержек). Средняя величина в данном случае показывает величину соответствующего показателя в расчете на единицу выпуска, предельная - прирост соответствующего показателя в расчете на единицу прироста выпуска. Так, если средняя выручка превышает средние издержки, то фирма получает прибыль и производить продукцию выгодно. Если при этом предельная выручка превышает предельные издержки, то фирме выгодно расширять производство, увеличивая объем прибыли. Соответственно, если средние издержки превышают среднюю выручку, то фирма терпит убытки, а если предельные издержки превышают предельную выручку, то объем производства нужно сократить. Далее мы рассмотрим формально роль абсолютных (суммарных) и относительных (средних и предельных величин) в экономическом анализе, а также свойства и соотношения этих величин.

6.2. Определение и геометрическая интерпретация суммарных, средних и предельных величин

Суммарная величина ($F(x)$). Под суммарной величиной мы будем понимать любую функцию независимой переменной $F(x)$. Как правило, в экономике под суммарными понимаются абсолютные величины, но, вообще говоря, формальное понятие суммарной величины является относительным (то есть любая величина может рассматриваться как суммарная по отношению к другим, своим предельным и средним величинам). В экономике в роли суммарных величин выступают: доход (выручка) или издержки как функции объема выпуска ($R(Q)$ или $C(Q)$), объем выпуска как функция от количества переменного ресурса, например труда, - $Q(L)$, полезность как функция количества потребляемого блага $U(x)$ и другие экономические показатели. Любая из перечисленных функций может быть задана в виде формулы, например, $F(x)=ax^2 - bx$; графика, например, показанного на рис. 6.1, и т.д.

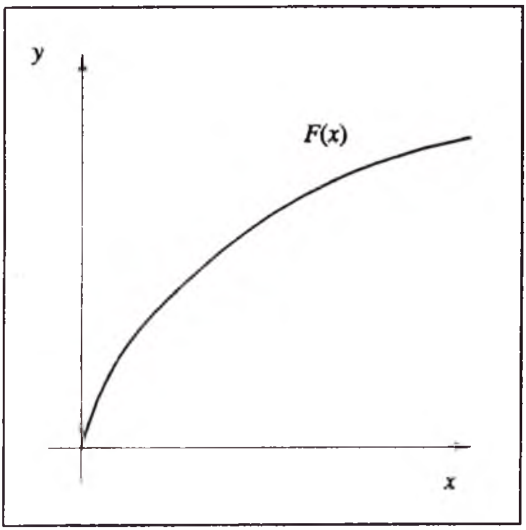


Рис. 6.1

Средняя величина ($AF(x)$) определяется как отношение суммарной величины к независимой переменной $AF(x) = \frac{F(x)}{x}$. Буква A - сокращение от Average (средняя). Средняя величина может обозначаться также $\bar{F} \equiv AF(x)$. Примеры средних величин в экономике: среднедушевой объем потребления, средняя фондоотдача, средняя выручка (доход) $AR = \frac{R(Q)}{Q}$, средние издержки $AC = \frac{C(Q)}{Q}$, средний продукт труда $AQ_L = \frac{Q(L)}{L}$ и т.д.

Средняя величина, как функция независимой переменной, также может задаваться в формульном или графическом виде.

Маржинальная (предельная) величина ($MF(x)$) определяется как производная суммарной величины $F(x)$ по независимой переменной x: $MF(x) = F'(x) \equiv \lim\limits_{\Delta x \to 0}\frac{\Delta F(x)}{\Delta x}$ в случае, когда независимая переменная меняется непрерывно. Если суммарная величина меняется дискретно, то под маржинальной (предельной) величиной понимают отношение изменения $\Delta F(x)$ суммарной величины $F(x)$ к вызвавшему это изменение изменению (приращению) Δx независимой переменной x: $MF(x) = \frac{\Delta F}{\Delta x}$. В этом случае маржинальную (предельную) величину можно интерпретировать как изменение суммарной величины, вызванное увеличением независимой переменной на единицу (в соответствующем масштабе). Примеры предельных величин в экономике: предельная выручка (доход) $MR = R'(Q)$ или $\frac{\Delta R}{\Delta Q}$, предельные издержки $MC = C(Q)$ или $\frac{\Delta C}{\Delta Q}$, предельный продукт труда $MQ_L = Q'(L)$ или $\frac{\Delta Q}{\Delta L}$, предельная полезность $MU_x = U'(x)$ или $\frac{\Delta U}{\Delta x}$ и т.д. Предельная величина, как и все предыдущие, может задаваться формулой или в графическом виде.

Встречаясь с этими величинами в экономике, часто приходится использовать соотношения между ними (например, между суммарными, средними и предельными издержками) и решать задачи на нахождение по одной из этих величин двух других (например, среднего и предельного дохода по суммарному доходу).

6.3. Соотношения между суммарными, средними и маржинальными (предельными) величинами. Задачи нахождения по одной из этих величин двух других. Формальный и графический анализ

Нахождение средней величины по суммарной. Формальный и графический анализ. Формально, эта задача решается с помощью определения средней величины: $AF(x) = \frac{F(x)}{x}$; например, если $F(x) = ax - bx^3$, то $AF(x) = \frac{ax - bx^3}{x} = a - bx^2$. Для графического решения этой задачи (когда суммарная величина задана в виде графика) необходимо провести вектор, соединяющий начало координат с точкой графика функции, имеющей координаты $(x, F(x))$.

Тангенс угла наклона этого вектора, равный отношению противолежащего (углу β) катета прямоугольного треугольника $F(x)$ к прилежащему x: $\operatorname{tg}\beta = \frac{F(x)}{x}$, будет (по определению) численно равен средней величине $AF(x) = \operatorname{tg}\beta(x)$ при любом значении независимой переменной x, отличном от нуля.

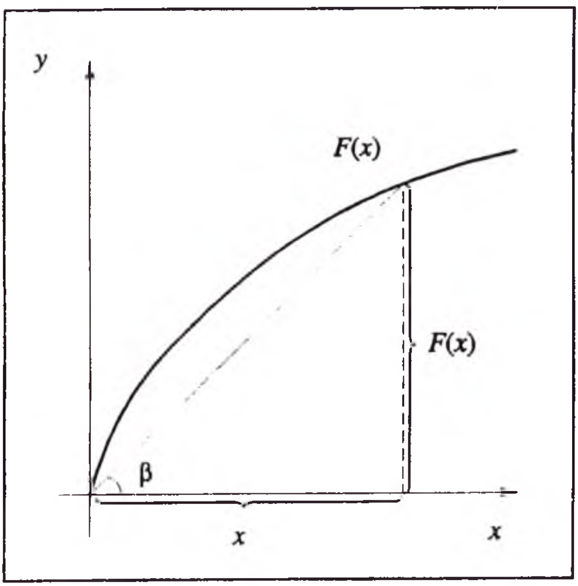


Рис. 6.2

Геометрическая интерпретация изменения средней величины. При изменении независимой переменной x угол наклона вектора в также изменяется. Увеличение этого угла с увеличением x свидетельствует о возрастании средней величины, а уменьшение - об убывании. В частности, для приведённого на рис. 6.2 графика средняя величина убывает, и её график изображен на рис. 6.3.

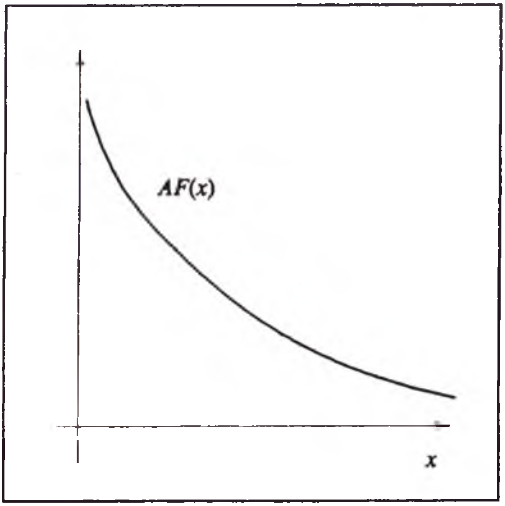


Рис. 6.3

Нахождение суммарной величины по средней (обратная задача). Формально обратная задача решается так же, как и прямая, с помощью определения, из которого находим $F(x) = x\,AF(x)$ (см.рис.3).

Если средняя величина задана в виде графика, представленного на рис. 6.4, то суммарную величину при данном значении независимой переменной x можно определить как площадь прямоугольника с вершинами в начале координат и точке графика средней величины, имеющей координаты (x, $AF(x)$), и сторонами x и $AF(x)$. Определяя характер изменения площади, мы можем построить график суммарной величины. Однако, на практике, при качественном построении графиков, удобнее применить “метод подбора”, т.е. подобрать такую функцию $F(x)$, чтобы наклон прямой, соединяющей точки ее графика с началом координат, изменялся в соответствии с заданным характером изменения средней величины $AF(x)$.

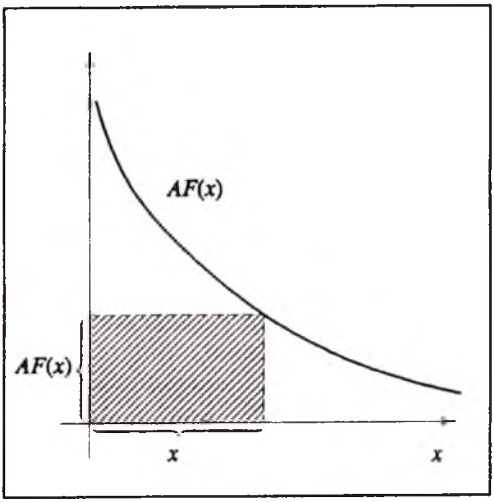


Рис. 6.4

Нахождение маржинальной (предельной) величины по суммарной (для непрерывного случая). Формальный и графический анализ. Формально эта задача решается с помощью определения предельной величины $MF(x) = F'(x)$; например, если $F(x) = ax - bx^3$, то $MF(x) = (ax - bx^3)' = a - 3bx^2$. Для графического решения этой задачи (когда суммарная величина задана в виде графика) необходимо через точку графика суммарной величины, имеющую координаты $(x, F(x))$, провести касательную к графику. Тангенс угла наклона касательной к графику суммарной величины в произвольной точке x согласно геометрической интерпретации производной будет численно равен производной суммарной величины, а следовательно, являться предельной величиной $MF(x) = F'(x) = \text{tg}\alpha_{\text{касательной}}(x)$ (см. рис. 6.5).

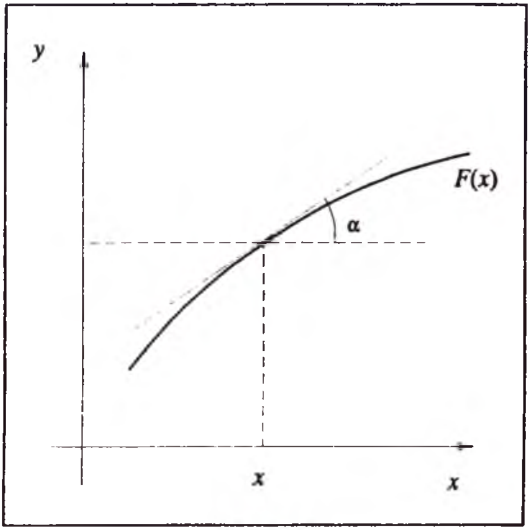


Рис. 6.5

Геометрическая интерпретация изменения маржинальной (предельной) величины. При изменении независимой переменной x угол наклона касательной б также изменяется. Увеличение этого угла с увеличением x свидетельствует о возрастании предельной величины, а уменьшение - об убывании. В частности, для приведённого ниже графика суммарной величины предельная величина убывает, и её график имеет вид:

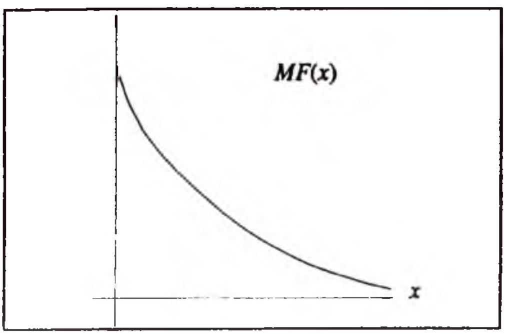


Рис. 6.6

Нахождение суммарной величины по маржинальной (предельной) (обратная задача). Формально обратная задача означает нахождение функции $F(x)$, производная которой $F'(x) \equiv MF(x)$ известна. Для решения этой задачи служит операция интегрирования, обратная операции дифференцирования. Функция $F(x)$ называется первообразной для функции $MF(x)$ и находится с помощью неопределенного интеграла $F(x) = \int MF(x)dx$. Например, если $MX(x) = a - \frac{bx^2}{3}$, то

$$F(x) = \int MF(x)dx = \int \frac{a - bx^2}{3}dx = ax - bx^3 + C,$$ где C - произвольная постоянная. Здесь мы воспользовались известной формулой интегрирования степенной функции $\int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1} + C$ $(\alpha \neq -1)$. В большинстве практических задач применение операции интегрирования часто можно заменить применением "метода подбора" - т.е. подобрать такую функцию $F(x)$, что ее производная будет равна данной в задаче функции $MF(x)$. Этот прием особенно удобно применять для степенных функций и многочленов, когда операция дифференцирования вызывает понижение степени на единицу, а операция интегрирования - повышение степени на единицу.

Если предельная величина задана в виде графика, то, согласно геометрической интерпретации неопределенного интеграла, площадь под графиком функции предельной величины в диапазоне изменения независимой переменной от нуля до x будет равна суммарной величине минус некоторая постоянная C или $F(x) = S(x) + C$.

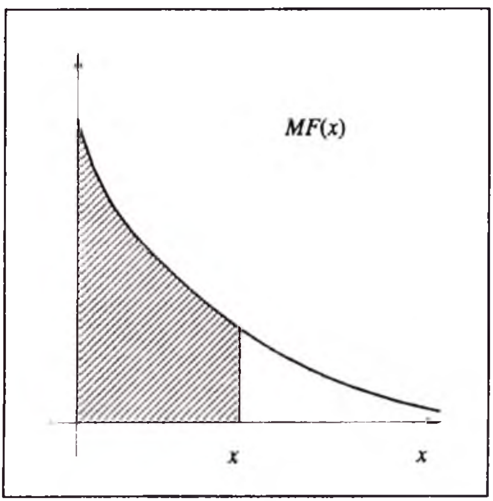


Рис. 6.7

Если при $x \to 0$ площадь $S(x) \to 0$, то константу можно найти как значение суммарной величины при $x = 0$. Например, при нахождении суммарного дохода $R(Q) = pQ$, чаще всего $R(0) = 0$ и $C = 0$; при нахождении суммарных издержек $C(Q)$ постоянная $C = C(0)$ имеет смысл фиксированных (постоянных) издержек.

Соотношения между средними (AF) и маржинальными (предельными) (MF) величинами. Задача нахождения по одной из этих величин другой может быть сведена к одной из предыдущих задач, если, предварительно, мы найдём суммарную величину. Например, если дана средняя величина $AF(x)$, то суммарная величина $F(x) = x \cdot AF(x)$, а предельная $MF(x) = F'(x) = (x \cdot AF(x))' = AF(x) + x \cdot AF'(x)$. Аналогично, можно выразить среднюю величину через суммарную $AF(x) = \frac{1}{x}\int MF(x)dx$. Первое из этих соотношений, а именно соотношение $MF(x) = AF(x) + x \cdot AF'(x)$, имеет простую интерпретацию. В точке экстремума функции $AF(x)$ ее производная $AF'(x) = 0$, и, следовательно, предельная величина совпадает со средней в точке экстремума последней. Предположим, что независимая переменная может принимать только положительные значения ($x > 0$), тогда:

а) в области возрастания функции $AF(x)$ её производная $AF'(x) > 0$, и $MF(x) > AF(x)$ (предельная величина больше средней); б) в области убывания функции $AF(x)$ её производная $AF'(x) < 0$ и $MF(x) < AF(x)$ (предельная величина меньше средней). Таким образом, график предельной величины лежит выше графика средней величины в области возрастания последнего, ниже - в области убывания, и проходит через точку экстремума графика средней величины.

Примеры соотношения между графиками средних и предельных издержек (AC и MC), а также между графиками среднего и предельного продуктов труда (AP_L и MP_L), приведены ниже.

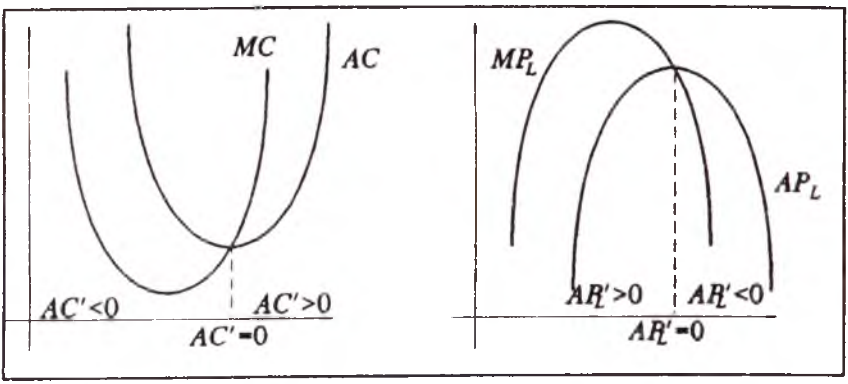


Рис. 6.8 Рис. 6.9

Дискретный случай. Если независимая переменная x может принимать только дискретные значения (например, объём выпуска автомобилей фирмой, количество нанимаемых ею рабочих и т.д.), то все полученные выше соотношения сохраняют свой вид при следующих условиях: а) производная функция $F'(x)$ заменяется на отношение $\frac{\Delta F}{\Delta x}$; б) интеграл $\int MF(x)dx$ заменяется на конечную сумму $\sum_x MF(x)$; в) касательная к графику функции $F(x)$ заменяется на прямую линию, проходящую через две точки с координатами $(x, F(x))$ и $(x+\Delta x,\ F(x+\Delta x))$.

Соотношение между средними и предельными величинами в дискретном случае имеет простую интерпретацию. Представим себе ученика-"хорошиста", который получает одни четвёрки. Каждую последующую оценку можно интерпретировать как предельную оценку, а средний балл - как среднюю оценку. Если этот ученик решит стать "отличником" и будет получать в дальнейшем одни пятёрки (предельная оценка выше средней), то его средняя оценка будет постепенно повышаться. Если ученик, наоборот, разленится и превратится в "троечника", начав получать одни тройки (предельная оценка станет меньше средней), то его средняя оценка будет понижаться.

6.4. Функции суммарного, среднего и предельного дохода и издержек

В качестве примера применения соотношений между предельными, средними и суммарными величинами рассмотрим экономические показатели, характеризующие работу фирмы: Q - объем выпуска, p - цена, $R = p(Q)\cdot Q$ - доход (выручка), C - издержки, $\Pi = R - C$ - прибыль - для двух типов рыночной структуры: совершенной конкуренции и монополии.

Совершенная конкуренция. В этом случае цена на продукцию фирмы не зависит от объёма производства данной фирмы, а определяется рынком и постоянна. Цена $p(Q) = p$ и, следовательно, $R(q) = p\cdot Q$. Доход является линейной функцией объёма выпуска. Для типичной функции издержек (растущих быстрее чем доход при малых объёмах выпуска) графики дохода, издержек и прибыли показаны на рис. 10.

По ним можно построить графики средних и предельных величин. Так как $MR = (pQ)' = p = \frac{pQ}{Q} = AR$, то графики среднего и предельного дохода имеют вид прямой, параллельной оси Q. График средних издержек совпадает с графиком среднего дохода при

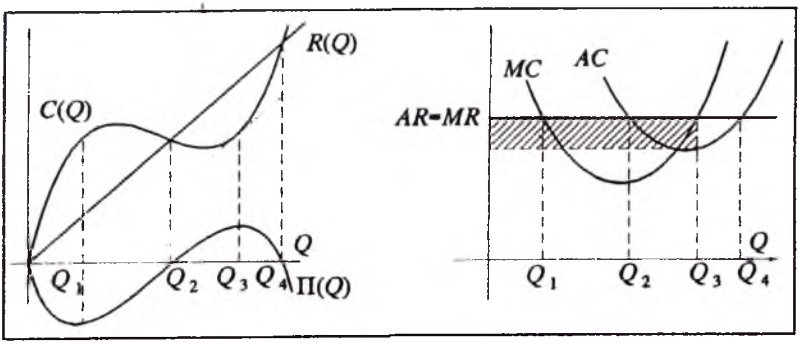


Рис. 6.10 *Рис. 6.11*

объёмах выпуска Q_2 и Q_4 (так как в этих точках значения функций $C(Q)$ и $R(Q)$ совпадают), лежит выше него при $Q < Q_2$ и $Q > Q_4$ (из $C(Q) > R(Q) \Rightarrow AC(Q) = \frac{C(Q)}{Q} > \frac{R(Q)}{Q} = MC(Q)$) и ниже - при $Q_2 < Q < Q_4$. В точке, лежащей чуть ниже Q_3, средние издержки минимальны. Эту точку можно найти, проводя из начала координат прямую, касающуюся графика $C(Q)$. График предельных издержек можно построить, анализируя изменение наклона касательной к графику $C(Q)$. В точках Q_1 и Q_3 касательная к графику $C(Q)$ параллельна графику дохода $R(Q)$. Следовательно, в этих точках предельные издержки совпадают с предельным доходом и имеет место минимум прибыли (максимум убытков) в точке Q_1 и максимум прибыли в точке Q_3 ($П' = R' - C' = MR - MC = 0$, ибо, как видно из рис. 6.10, прибыль положительна при объёме выпуска $Q_2 < Q < Q_4$ и отрицательна при $Q < Q_2$ и $Q > Q_4$). Величину прибыли при оптимальном объёме выпуска (Q_3) можно найти как площадь заштрихованного прямоугольника по графикам средних издержек и среднего дохода. (Вершины прямоугольника находятся в точках с координатами: ((Q_3, p), (Q_3, $AC(Q_3)$), (0, $AC(Q_3)$), (0, p).)

Монополия. В случае монополии фирма сама выбирает цену, исходя из кривой спроса $p(Q)$ на её продукцию. Поскольку $p(Q)$ - убывающая функция, $p'(Q)<0$. При той же функции издержек, что и в предыдущем случае, графики суммарных, средних и предельных показателей показаны на рис. 6.12, 6.13. При этом графики суммарных, средних и предельных издержек имеют тот же вид, что и предыдущем случае.

График среднего дохода $AR = \frac{p(Q)\cdot Q}{Q} = p(Q)$ совпадает с графиком функции спроса и пересекает график средних издержек в точках Q_2 и Q_4 (где $R(Q) = C(Q)$). График предельного дохода лежит

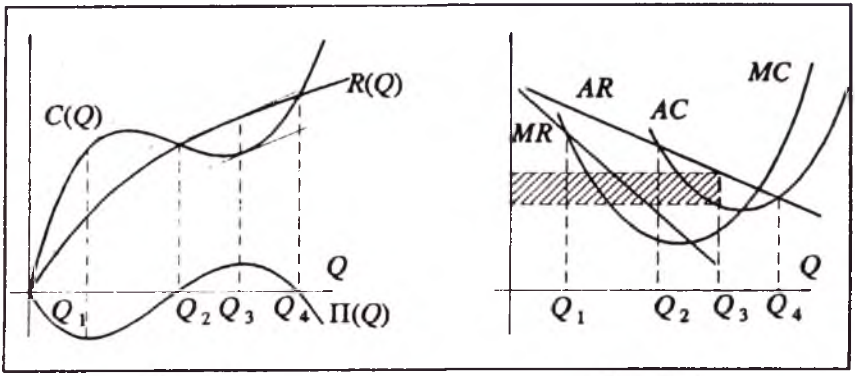


Рис. 6.12 Рис. 6.13

ниже графика среднего дохода при любых объёмах выпуска, так как $MR = R'(Q) = (p(Q)\ Q)' = p(Q) + Qp'(Q) = AR + Qp'(Q) < AR$, (поскольку $p'(Q) < 0$), и пересекает график предельных издержек в точках Q_1 и Q_3, в которых касательные к графикам дохода и издержек имеют одинаковый наклон. При этих объёмах выпуска прибыль, как и в предыдущем случае, принимает минимальное и максимальное значения соответственно. Это обусловлено тем, что необходимое условие максимума прибыли по-прежнему записывается как $\Pi' = R' - C' = MR - MC = 0$, и в оптимальной точке предельный доход обязательно равен предельным издержкам. Аналогично предыдущему случаю, прибыль на графиках средних и предельных величин также можно определить как площадь заштрихованного прямоугольника, построенного между графиками среднего дохода и средних издержек (вершины прямоугольника находятся в точках: $(Q_3,AR(Q_3))$, $(Q_3,AC(Q_3))$, $(0,AC(Q_3))$, $(0,AR(Q_3))$).

Итак, при определении оптимального объема производства фирмы, если известны ее функции суммарного дохода и издержек $R(Q)$ и $C(Q)$ (предполагается, что эти функции дифференцируемы), средние и предельные показатели могут быть использованы следующим образом.

Вначале находятся точки, в которых величина предельного дохода равна величине предельных издержек: $MR(Q)=MC(Q)$. Если таких точек нет, то фирме либо невыгодно производить вообще (при $R(Q)<C(Q)$), либо выгодно сколь угодно наращивать объем производства (при $R(Q)>C(Q)$).

В найденных точках может достигаться максимум прибыли, максимум убытка, минимум прибыли, минимум убытка, либо ничего из перечисленного. Поэтому далее среди этих точек находятся те, в которых функция прибыли $\Pi(Q)= R(Q)-C(Q)$ достигает максимума (ее производная меняет знак с плюса на минус). Это точки максимума прибыли или минимума убытка. Наконец, нужно выбрать точки (точку), где величина прибыли положительна. Признаком этого может быть превышение среднего дохода над средними издержками: $AR(Q)>MR(Q)$. Если такая точка найдена, то она является точкой (локального) максимума прибыли фирмы.

Вопросы к главе 6

1. Что такое абсолютные и относительные величины в экономике? Приведите примеры.
2. Что такое суммарная величина? Приведите примеры суммарных величин в экономике.
3. Приведите примеры графического представления суммарных величин в экономике.
4. Дайте определение средней величины. Приведите примеры средних величин в экономике.
5. Приведите примеры графического представления средних величин в экономике.
6. Что такое предельная величина? Приведите примеры предельных величин в экономике. Каково различие в определениях средней величины в дискретном и в непрерывном случаях? Приведите практические примеры.
7. Приведите примеры графического представления предельных величин в экономике.
8. Суммарные издержки имеют вид $C(Q) = 2Q - 3Q^2 + Q^3$. Найдите средние и предельные издержки.
9. Пусть график суммарного дохода представляет из себя возрастающую функцию. Что Вы можете сказать о графиках предельного и среднего дохода? Какая дополнительная информация Вам потребуется?
10. Пусть график среднего дохода представляет из себя убывающую функцию. Что Вы можете сказать о графиках суммарного и предельного дохода? Какая дополнительная информация Вам потребуется?
11. Каково соотношение между средними и предельными издержками, если и те и другие возрастают? Достаточно ли этой информации для ответа на вопрос?
12. Что можно сказать о зависимости дохода предприятия от объема производства, если известно, что при любом объеме производства предельный доход совпадает со средним?
13. Что можно сказать о зависимости прибыли предприятия от объема производства, если известно, что при любом объеме производства, большем некоторого значения Q_0, предельный доход меньше предельных издержек?
14. Какими двумя способами Вы могли бы определить прибыль фирмы по известным графикам предельных и средних издержек и дохода?
15. Как использовать суммарные, средние и предельные величины для определения оптимального объема производства фирмы?

ГЛАВА 7
ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ И ИХ ЭКСТРЕМУМЫ

7.1. Функции двух переменных и их множества (линии) уровня

В функции *двух* переменных независимых переменных *две*, а не *одна* как в случае функции *одной* переменной:

$y = f(x)$ - функция *одной* переменной x;

$y = f(x_1, x_2)$ - функция *двух* переменных x_1 и x_2.

Переменные x_1 и x_2 меняются независимо друг от друга.

Если независимых переменных $x_1,\ldots,x_n$ n (штук), имеем функцию n переменных $y=f(x_1,\ldots,x_n)$.

В экономических приложениях математики широко используются линейные и нелинейные функции двух переменных и n переменных.

Пример 1.1. Функции $y=a_0+a_1x_1+a_2x_2$, $y=a_0+a_1x_1+\ldots+a_nx_n$ - линейные функции двух и n переменных. Функции $y=x_1^2+x_2^2$, $y=x_1^2+\ldots+x_n^2$, $y=x_1^2-x_2^2$, $y=a_0x_1^{a_1}x_2^{a_2}$, $y=a_0x_1^{a_1}x_2^{a_2}\ldots x_n^{a_n}$, $y=\min(x_1/b_1,x_2/b_2)$, $y=\min(x_1/b_1,x_2/b_2,\ldots,x_n/b_n)$ являются нелинейными.

График Γ функции двух переменных определяется аналогично графику Γ функции одной переменной.

Графиком функции f двух переменных x_1 и x_2 называется множество точек (x_1, x_2, y) трехмерного пространства таких, что $y = f(x_1, x_2)$, т.е. множество точек $(x_1, x_2, f(x_1, x_2))$. Обычно в экономических приложениях $x_1 > 0$, $x_2 > 0$. График Γ функции двух переменных можно наглядно представить в виде двумерной поверхности в трехмерном пространстве. Для функции трех и более переменных понятие графика Γ определяется аналогично как множество точек $(n + 1)$-мерного пространства $(x_1, \ldots, x_n, f(x_1, \ldots, x_n))$. Однако при $n > 2$ график Γ уже не имеет наглядного геометрического представления (как это имеет место при $n = 1$ и $n = 2$) (компьютерная графика позволяет смотреть проекции).

Пример 1.2. Построим график Γ функции $y=x_1^{1/2}x_2^{1/2}$ при $x_1\geq 0$, $x_2\geq 0$ (эта функция представляет собой конкретный пример производственной функции Кобба-Дугласа (ПФ КД), когда $a_1=a_2=1/2$, $a_0=1$) (см. рис. 7.1).

Очевидно, при $x_1 \geq 0$, $x_2 \geq 0$ график Γ есть коническая поверхность, образующие которой - лучи, выходящие из точки O, а направляющая есть линия H (см. рис. 7.1). В вертикальной плоскости $x_1+x_2=1$ линия H имеет уравнение $y=x_1^{1/2}(1-x_1)^{1/2}$.

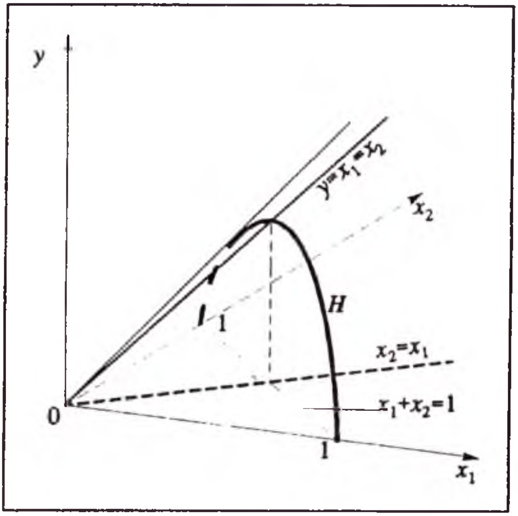


Рис. 7.1

Пример 1.3. Построить график Γ функции $y=x_1^{1/4}x_2^{1/4}$ при $x_1 \geq 0$, $x_2 \geq 0$ (здесь $a_1=a_2=1/4, a_0=1$) (см. рис. 7.2).

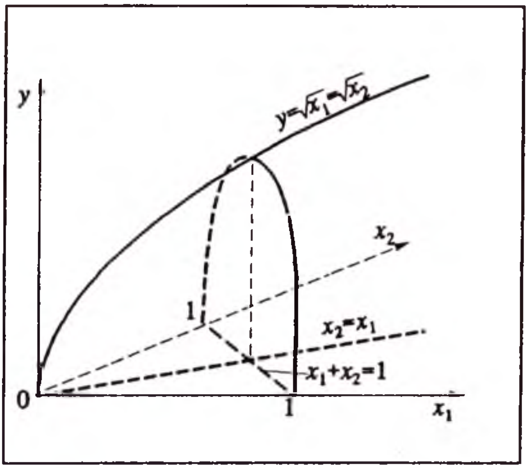


Рис. 7.2

Пример 1.4. Построить самостоятельно график Γ функции $y=x_1x_2$ при $x_1\geq 0$, $x_2\geq 0$ (здесь $a_1=a_2=1, a_0=1$).

В экономических приложениях широко используются понятия выпуклого множества и выпуклой функции двух и нескольких переменных. Сначала приведём определение для случая, когда $n=2$.

Определение 1.1.

Множество называется выпуклым, если оно вместе с двумя любыми своими точками содержит отрезок, их соединяющий (см. рис. 7.3а).

Множество, которое не является выпуклым, называется невыпуклым (см. рис. 7.3б). Приведённое определение выглядит одинаково для случая двух переменных и для случая n (нескольких) переменных.

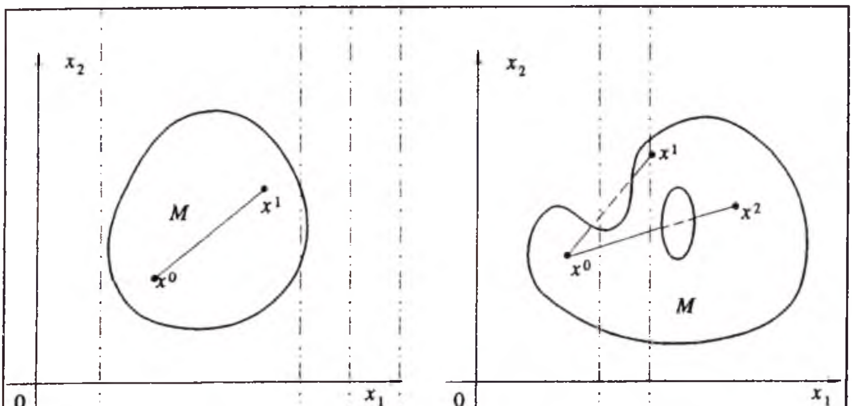


Рис. 7.3а *Рис. 7.3б*

Наглядно понятие выпуклого множества можно пояснить так: выпуклое множество - это множество, которое не имеет вмятин и дыр. На рис. 7.3а изображено выпуклое множество, на рис. 7.3б представлено невыпуклое множество M, которое имеет одну вмятину и одну дыру.

Определение 1.2.

Функция $f(x)$, определённая на выпуклом множестве M, называется выпуклой вниз (вогнутой вверх), если для любых двух точек x^0 и x^1 из множества M и для любого числа t $0<t\leq 1$ справедливо неравенство $f((1-t)x^0+tx^1)$ $(1-t)f(x^0)+tf(x^1)$. Например, функции $f(x)=a_1x_1+a_2x_2+a_3$, $f(x)=x_1^2+x_2^2$ выпуклы вниз на всём пространстве E_2.

График Γ_f выпуклой вниз функции $f(x)$ расположен ниже (точнее не выше) любой своей хорды (см. рис. 7.4).

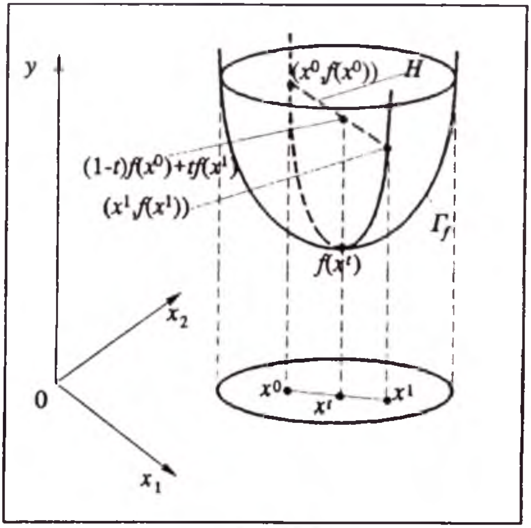


Рис. 7.4

Определение 1.3.

Функция $g(x)$, определённая на выпуклом множестве M, называется выпуклой вверх (вогнутой вниз), если функция $g(x)=-f(x)$, где функция $f(x)$ выпукла вниз. Например, функции $y=a_1x_1+a_2x_2+a_3$, $y=a_0x_1^{\alpha_1}x_2^{\alpha_2}$ $(0<\alpha_1+\alpha_2<1)$ выпуклы вверх.

Термины "выпуклый вниз" ("вогнутый вверх"), "выпуклый вверх" ("вогнутый вниз") применяются также к графикам соответствующих функций.

Для случая $n>2$ приведённые определения функции выпуклой вниз и выпуклой вверх переписываются с незначительными корректировками.

Множеством (чаще говорят - *линией) уровня* q (q - число, в экономических приложениях $q\geq 0$) функции $y=f(x_1, x_2)$ называется множество (совокупность) всех пар (x_1, x_2) такое, что $f(x_1, x_2) = q$, т.е. во всех точках (x_1, x_2), принадлежащих множеству уровня q, частное значение функции $y = f(x_1, x_2)$ одно и то же и равно q. Множество уровня q функции $y=f(x_1, x_2)$ обозначается символом l_q. На рис. 7.5 наглядно иллюстрируется это важное математическое понятие. Горизонтальная плоскость P пересекается с графиком Γ по плоской горизонтальной линии L_q, которая вся "зависает" над плоскостью Ox_1x_2 на высоте q. Проектируя L_q на плоскость Ox_1x_2, получаем линию l_q, которая и есть множество уровня q функции $y = f(x_1, x_2)$. Специально отметим, что все точки линии L_q принадлежат графику $Г$.

Множество всех множеств (линий) уровня функции $y = f(x_1, x_2)$ называется *картой* линий уровня функции $y = f(x_1, x_2)$. По карте можно получить довольно точное представление о характере графика $Г$ функции $f(x_1, x_2)$. Фрагмент карты линий уровня функции $y = f(x_1, x_2)$, график $Г$ которой представлен на рис. 7.5, изображен на рис. 7.4.

График $Г$ имеет вид "горки" (см. рис. 7.3), поэтому линия соответствует уровню q_2, который больше уровня q, т.е. $q_2 > q$. Аналогично, $q_1 < q$. Таким образом, если график $Г$ имеет вид "горки", то линии уровня расположенной северо-восточнее линии уровня l_q (или l_{q1}), соответствует и больший уровень q_2, т.е. $q_2>q$ ($q_2>q_1$). Верно и обратное (см. рис. 7.5).

Если линия (или линия l_q или линия l_{q2}) выглядит так, как на рис. 6, то говорят, что эта линия *выпукла* к точке О.

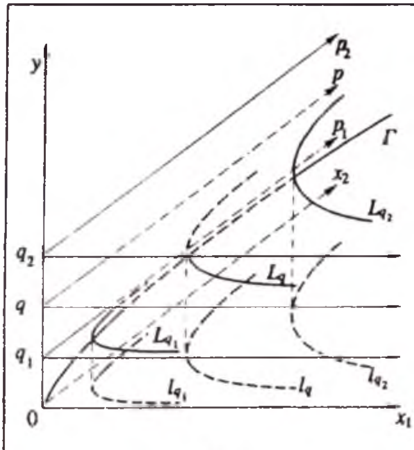


Рис. 7.5

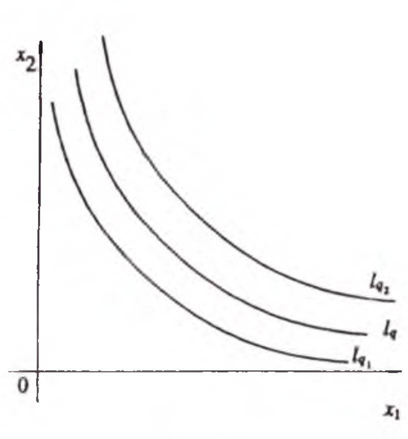


Рис. 7.6

7.2. Частные производные, градиент и дифференциал

Определение 2.1.

Пусть $y = f(x_1, x_2)$ - функция двух переменных. (Первая) производная функции $f(x_1, x_2)$ по переменной x_1 при фиксированной второй переменной x_2 называется (первой) *частной производной* функции $f(x_1, x_2)$ по переменной x_1, что символически записывается так:

$$\frac{\partial f(x_1, x_2)}{\partial x_1}, \text{ или } \frac{\partial y(x_1, x_2)}{\partial x_1}, \text{ или просто } \frac{\partial y}{\partial x_1}.$$

Аналогично определяется (первая) частная производная функции $f(x_1, x_2)$ по переменной x_2:

$$\frac{\partial f(x_1,x_2)}{\partial x_2}, \text{ или } \frac{\partial y(x_1,x_2)}{\partial x_2}, \text{ или просто } \frac{\partial y}{\partial x_2}.$$

Обратим внимание, что в символике частных производных используются круглые ∂, а не прямые d. В случае (первой) частной производной $\dfrac{\partial f(x_1,\ldots,x_i,\ldots,x_n)}{\partial x_i}$ по переменной x_i функции $f(x_1, \ldots, x_i, \ldots, x_n)$ n переменных роль постоянных играют все переменные $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$, кроме переменной x_i.

Первая частная производная по переменной x_1 представляет собой, вообще говоря, новую функцию двух (нескольких) переменных. Если нет специальной оговорки, везде в данном учебном пособии предполагается, что частные производные принимают только конечные значения, т.е. речь идёт только о конечных частных производных.

Если в точке (x_1^0, x_2^0) значение (первой) частной производной функции $f(x_1,x_2)$ по переменной x_1 (x_2) положительно, т.е. если $\dfrac{\partial f(x_1^0,x_2^0)}{\partial x_1} > 0$, $\left(\dfrac{\partial f(x_1^0,x_2^0)}{\partial x_2} > 0\right)$, то при *малом* росте переменной x_1 (x_2) относительно x_1^0 (x_2^0) при фиксированной переменной x_2^0 (x_1^0) значение y функции $f(x_1, x_2)$ растет, т.е. из того, что (первая) частная производная по переменной x_1 (x_2) положительная, следует свойство (локальной) монотонности функции $f(x_1,x_2)$ по переменной x_1 (x_2) (см. раздел 1).

Пример 2.1. Имеем $y = a_0 + a_1x_1 + a_2x_2$, тогда $\dfrac{\partial y}{\partial x_1} = a_1$, $\dfrac{\partial y}{\partial x_2} = a_2$.

При нахождении частной производной $\dfrac{\partial y}{\partial x_1}$ слагаемое a_2x_2 фиксировано, т.е. играет роль постоянной, как и слагаемое a_0, поэтому производная по x_1 суммы (“хвоста”) $(a_0+a_2x_2)$ равна нулю. Аналогично поясняется ответ $\dfrac{\partial y}{\partial x_2} = a_2$. Также по аналогии, в случае $y = a_0 + a_1x_1 + \ldots + a_ix_i + \ldots + a_nx_n$ имеем n штук (первых) частных производных $\dfrac{\partial y}{\partial x_i} = a_i$, $i = 1, \ldots, n$.

Пример 2.2. Производная степенной функции $y=b_0x^{a_1}$ одной переменной x равна $\dfrac{dy}{dx} = b_0a_1x^{a_1-1}$. (Первая) частная производная функции $a_0x_1^{a_1}x_2^{a_2}$ по переменной x_1 равна $\frac{\partial y}{\partial x_1} = a_0a_1x_1^{a_1-1}x_2^{a_2}$.

Здесь роль b_0 играет произведение $a_0x_2^{a_2}$. Аналогично $\frac{\partial y}{\partial x_2} = a_0a_2x_1^{a_1}x_2^{a_2-1}$.

Также по аналогии в случае $y=a_0x_1^{a_1}...x_i^{a_i}...x_n^{a_n}$ имеем n штук (первых) частных производных $\frac{\partial y}{\partial x_i} = a_0a_ix_1^{a_1}... x_i^{a_i-1} ... x_n^{a_n}$, $i = 1, ..., n$.

Определение 2.2.

Упорядоченная *пара* (первых) частных производных $\left(\frac{\partial f(x_1,x_2)}{\partial x_1}, \frac{\partial f(x_1,x_2)}{\partial x_2}\right)$ или $\left(\frac{\partial y(x_1,x_2)}{\partial x_1}, \frac{\partial y(x_1,x_2)}{\partial x_2}\right)$ функции $y = f(x_1, x_2)$ *двух* переменных x_1 и x_2 обозначается символом grad $f(x_1, x_2)$ (или $f'(x_1, x_2)$ или grad $y(x_1, x_2)$) и называется *градиентом* функции $y = f(x_1, x_2)$ двух переменных. Градиент функции *двух* переменных есть *двумерный* вектор, функции $f(x_1,...,x_n)$ n переменных - n-мерный вектор grad $f(x_1, ..., x_n) = \left(\frac{\partial f(x_1,\dots,x_n)}{\partial x_1},\dots,\frac{\partial f(x_1,\dots,x_n)}{\partial x_n}\right)$.

Градиент grad $f(x_1^0, x_2^0)$ функции $f(x_1, x_2)$ в точке (x_1^0, x_2^0) показывает направление самого быстрого *роста* функции $f(x_1, x_2)$ в точке (x_1^0, x_2^0).

Задача 2.1. Для функции $y=x_1^{1/2}x_2^{1/2}$ двух переменных x_1 и x_2

а) построить линию уровня проходящую через точку $(x_1^0, x_2^0) = (4, 1)$;

б) найти градиент $\left(\frac{\partial y(x_1^0,x_2^0)}{\partial x_1}, \frac{\partial y(x_1^0,x_2^0)}{\partial x_2}\right)$ в этой точке (4, 1);

в) построить этот градиент.

Решение задачи 2.1

а) Сначала найдем уровень q^0, который равен частному значению функции $y=x_1^{1/2}x_2^{1/2}$ в точке (4,1). Имеем: $q^0=(x_1^0)^{1/2}(x_2^0)^{1/2}=4^{1/2}1^{1/2}=2$. Построим на плоскости Ox_1x_2 линию уравнение которой имеет вид: $q_0=x_1^{1/2}x_2^{1/2}$, или $2=x_1^{1/2}x_2^{1/2}$ или $4=x_1x_2$, или, наконец $x_2 = \frac{4}{x_1}$ (см. рис. 7.7).

б) Имеем: $\frac{\partial y(x_1,x_2)}{\partial x_1} = ½\ x_1^{-1/2}\ x_2^{1/2}$, $\frac{\partial y(x_1,x_2)}{\partial x_2} = ½\ x_1^{1/2}\ x_2^{-1/2}$,

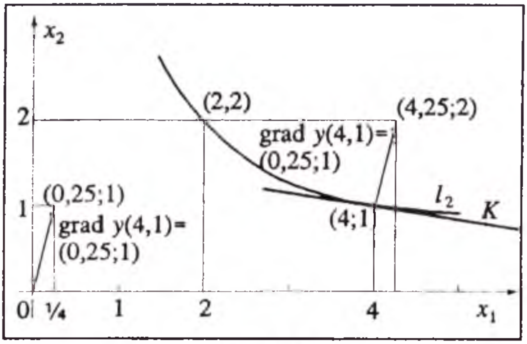


Рис. 7.7

$$\frac{\partial y(x_1^0, x_2^0)}{\partial x_1} = \frac{\partial y(1,4)}{\partial x_1} = \tfrac{1}{2}\ 4^{-1/2}\ 1^{1/2} = \tfrac{1}{2}\ \tfrac{1}{2} = \tfrac{1}{4},$$

$$\frac{\partial y(x_1^0, x_2^0)}{\partial x_2} = \frac{\partial y(1,4)}{\partial x_2} = \tfrac{1}{2}\ 4^{1/2}\ 1^{-1/2} = {}^2/_2 = 1.$$

Следовательно, grad $y(x_1^0, x_2^0) = \left(\frac{\partial y(x_1^0, x_2^0)}{\partial x_1}, \frac{\partial y(x_1^0, x_2^0)}{\partial x_2}\right) = \left(\frac{1}{4}, 1\right)$.

в) Строим градиент grad $y(x_1^0, x_2^0)=(1/4, 1)$ на плоскости Ox_1x_2, сначала выходящим из точки (0,0), а затем из точки (4,1) (см. рис. 7). Следует обратить внимание, что на рис. 7.7 grad $y(4,1)$ перпендикулярен (ортогонален) касательной K к линии (гиперболе) $l_2 = l_{q^0}$ в точке (4,1), т.е. ортогонален линии l_2, проходящей через точку (4,1). Этот частный факт есть иллюстрация общего случая: градиент grad $y(x_1^0, x_2^0)$ в точке (x_1^0, x_2^0) всегда ортогонален линии l_{q^0} уровня q_0, проходящей через точку (x_1^0, x_2^0).

Задача 2.2. Для функции $y=x_1^{1/2}x_2^{1/2}$ двух переменных x_1 и x_2:

а) построить (дополнив рис.7) линию уровня l_{q^1} проходящую через точку $(x_1^1, x_2^1)=(4,2)$;

б) найти градиент $\left(\frac{\partial y(x_1^1, x_2^1)}{\partial x_1}, \frac{\partial y(x_1^1, x_2^1)}{\partial x_2}\right)$ в точке (4,2);

в) построить этот градиент (дополнив рис. 7.7);

г) убедиться, что $q_1 > q_0 = 2$ и что линия l_{q^1} расположена “северо-восточнее” линии l_{q^0};

д) убедиться, что действительно, grad $y(4,1)$ показывает направление, в котором функция $y=x_1^{1/2}x_2^{1/2}$ растет.

Эту задачу предлагается решить самостоятельно.

Говорят, что уравнение $q = f(x_1, x_2)$ задает *неявную* функцию $x_2 = h(x_1)$ как функцию переменной x_1, ибо в уравнении $q = f(x_1, x_2)$ еще не выделена переменная x_2, как это имеет место в случае уравнения $x_2 = h(x_1)$. Аналогично можно говорить о неявной функции $x_1 = g(x_2)$ как функции переменной x_2.

Отметим, что если (первые) частные производные функции $f(x_1, x_2)$ непрерывны в точке (x_1^0, x_2^0) и в близких к ней точках и если для определенности $\dfrac{\partial f(x_1^0, x_2^0)}{\partial x_2} \neq 0$, то неявная функция $x_2 = h(x_1)$ существует при всех x_1 близких к x_1^0. Однако далеко не всегда на основании аналитического выражения $f(x_1, x_2)$ можно выписать аналитическое выражение для функции $x_2 = h(x_1)$.

Если линии уровня функции $y=f(x_1,x_2)$ являются нисходящими, т.е. линиями типа тех, что изображены на рис. 7.6 или на рис. 7.7, то для уравнения $q=f(x_1,x_2)$ неявная функция $x_2=h(x_1)$ (или $x_1=g(x_2)$) существует. Таким образом одна и та же нисходящая линия l (см. рис. 7.8) описывается уравнением $q = f(x_1, x_2)$, еще не разрешенным относительно переменной x_2 (или переменной x_1), и уравнением $x_2 = h(x_1)$ (или уравнением $x_1 = g(x_2)$), уже разрешенным относительно переменной x_2 (переменной x_1).

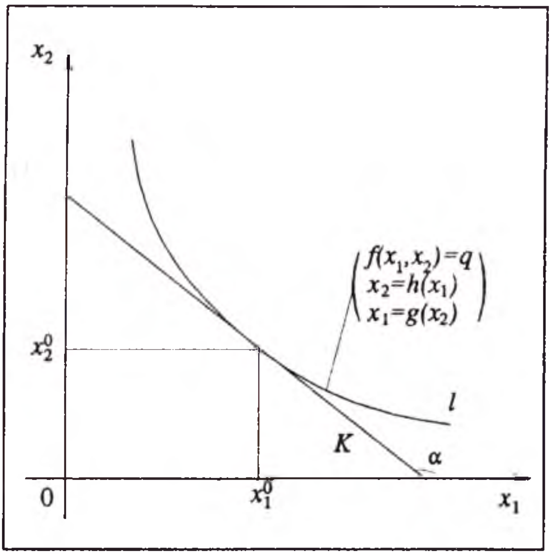


Рис. 7.8

Пример 2.3. Уравнение $2=x_1^{1/2}x_2^{1/2}$ можно переписать, выделив явно переменную x_2 (переменную x_1):

$$x_2=\frac{4}{x_1}\qquad \left[x_1=\frac{4}{x_2}\right]$$

Если (первые) частные производные $\dfrac{\partial f(x_1,x_2)}{\partial x_1}$ и $\dfrac{\partial f(x_1,x_2)}{\partial x_2}$ непрерывны в точке $(x_1^0,\ x_2^0)$ (см. рис. 7.8) и в близких к ней точках, то производную $h'(x_1^0)$ можно выписать, не используя явной формулы $x_2=h(x_1)$, следующим образом:

$$h'(x_1^0)=\frac{dh(x_1^0)}{dx_1}=-\left[\frac{\partial f(x_1^0,x_2^0)}{\partial x_1}\right]\Big/\left[\frac{\partial f(x_1^0,\ x_2^0)}{\partial x_2}\right].$$

Таким образом, $\mathrm{tg}\alpha$ (и, следовательно, наклон касательной K - см. рис. 7.8), равный $h'(x_1^0)$, может быть найден как отношение (первых) частных производных функции $f(x_1,\ x_2)$ в точке $(x_1^0,\ x_2^0)$, взятое со знаком минус, т.е. без использования явного выражения $h(x_1)$. Выписанная формула называется *производной неявной функции* $x_2=h(x_1)$. Эта формула играет важную роль в микроэкономическом анализе (в теории потребительского поведения и в теории фирмы). Производная неявной функции $x_1=g(x_2)$ выписывается аналогично (числитель и знаменатель меняются местами).

Пример 2.4. Пусть $x_1^{1/2}x_2^{1/2}=2$ и $(x_1^0,x_2^0)=(4,1)$. Имеем

$$\frac{\partial f(x_1,x_2)}{\partial x_1}=\tfrac{1}{2}\ x_1^{-1/2}\,x_2^{1/2},\ \frac{\partial f(x_1,x_2)}{\partial x_2}=\tfrac{1}{2}\ x_1^{1/2}\,x_2^{-1/2},\ h'(4)=-\frac{0{,}5(x_1^0)^{-\frac{1}{2}}(x_2^0)^{\frac{1}{2}}}{0{,}5(x_1^0)^{\frac{1}{2}}(x_2^0)^{-\frac{1}{2}}}$$

$$=-\frac{x_2^0}{x_1^0}=-\frac{1}{4}.$$

Определение 2.3.

По аналогии с (первым) дифференциалом $df(x)=f'(x)dx$ (или $dy(x)=y'(x)dx$ или $dy=y'dx$) функции $y=f(x)$ одной переменной x выражение

$$d_1f(x_1,x_2)=\frac{\partial f(x_1,x_2)}{\partial x_1}dx_1\ \left(\text{или } d_1y(x_1,x_2)=\frac{\partial y(x_1,x_2)}{\partial x_1}dx_1,\ \text{или } d_1y=\frac{\partial y}{\partial x_1}dx_1\right)$$

называется (первым) частным дифференциалом функции $y=f(x_1,\ x_2)$, соответствующим переменной x_1.

(Первый) частный дифференциал функции $y=f(x_1,\ x_2)$, соответствующий переменной x_2, имеет вид

$$d_2f(x_1,x_2)=\frac{\partial f(x_1,x_2)}{\partial x_2}dx_2\ \left(\text{или } d_2y(x_1,x_2)=\frac{\partial y(x_1,x_2)}{\partial x_2}dx_2,\ \text{или } d_2y=\frac{\partial y}{\partial x_2}dx_2\right)$$

Определение 2.4.

Сумма двух (первых) частных дифференциалов называется (первым) *полным дифференциалом* (символика: $df(x_1, x_2)$, $dy(x_1, x_2)$, dy) *функции* $y = f(x_1, x_2)$ двух переменных x_1 и x_2:

$$d_1 f(x_1,x_2) = \frac{\partial f(x_1,x_2)}{\partial x_1} dx_1 + \frac{\partial f(x_1,x_2)}{\partial x_2} dx_2 .$$

По аналогии для функции $y = f(x_1, \ldots , x_n)$ n переменных имеем следующее выражение для (первого) полного дифференциала:

$$df(x_1,\ldots,x_n) = \frac{\partial f(x_1,\ldots,x_n)}{\partial x_1} dx_1 + \ldots + \frac{\partial f(x_1,\ldots,x_n)}{\partial x_n} dx_n .$$

Пример 2.5. Для функции $y=x_1^{1/2}x_2^{1/2}$ имеем

$dy(x_1,x_2) = \dfrac{\partial y(x_1,x_2)}{\partial x_1} dx_1 + \dfrac{\partial y(x_1,x_2)}{\partial x_2} dx_2$, или $dy(x_1,x_2)=½(x_1^{-1/2}x_2^{1/2})dx_1 + ½(x_1^{1/2}x_2^{-1/2})dx_2$.

В точке $(x_1^0, x_2^0) = (4, 1)$ (первый) полный дифференциал имеет вид $dy(x_1^0,x_2^0) = \dfrac{\partial y(x_1^0,x_2^0)}{\partial x_1} dx_1 + \dfrac{\partial y(x_1^0,x_2^0)}{\partial x_2} dx_2$ или $dy(4,1)=¼dx_1+dx_2$.

7.3. Однородные функции

Определение 3.1.

Функция $y = f(x_1, x_2)$, определенная при $x_1 \geq 0$, $x_2 \geq 0$, называется *однородной функцией* степени p, если для любого числа $t > 0$ и любых $x_1 \geq 0$ и $x_2 \geq 0$ выполняется равенство

$$f(tx_1, tx_2) = t^p f(x_1, x_2).$$

Для функции $y = f(x_1, \ldots , x_n)$ n переменных определение аналогично

$$f(tx_1, \ldots , tx_n) = t^p f(x_1, \ldots , x_n).$$

Для однородных функций степени p (двух переменных) справедлива формула

$$\frac{\partial f(x_1,x_2)}{\partial x_1} x_1 + \frac{\partial f(x_1,x_2)}{\partial x_2} x_2 = pf(x_1,x_2).$$

Аналогичная формула имеет место и для однородной функции степени p n переменных

$$\frac{\partial f(x_1,\ldots,x_n)}{\partial x_1} x_1 + \ldots + \frac{\partial f(x_1,\ldots,x_n)}{\partial x_n} x_2 = pf(x_1,\ldots,x_2).$$

Приведенные для однородных функций степени p двух и n переменных формулы имеют место, если (первые) частные производные существуют и непрерывны (эти условия для многих производственных функций и функции полезности выполняются). Эти формулы называются *формулами Эйлера*, и утверждение об их справедливости - *теоремой Эйлера*. Формулы Эйлера существенно используются в микроэкономическом анализе.

Пример 3.1. Линейная функция вида $y = a_1x_1 + a_2x_2$ (она называется линейной формой) однородна первой степени, ибо $a_1(tx_1) + a_2(tx_2) = t(a_1x_1 + a_2x_2)$.

Пример 3.2. Квадратичная форма, т.е. функция вида $y = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$, однородна второй степени, ибо $a_{11}(tx_1)^2 + 2a_{12}(tx_1)(tx_2) + a_{22}(tx_2)^2 = t^2(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_{22}^2)$.

7.4. Элементы теории экстремума

Определение 4.1.

Двумерной δ-окрестностью точки (x_1^0,x_2^0) (символика: $U_2(\delta,x_1^0,x_2^0)$) называется множество точек (x_1^0,x_2^0), принадлежащих открытому кругу радиуса $\delta>0$ с центром в точке (x_1^0,x_2^0), т.е.

$$U_2(\delta,x_1^0,x_2^0) \overset{(def)}{=} \{(x_1,x_2)\,|\,(x_1-x_1^0)^2+(x_2-x_2^0)^2<\delta^2\}$$

(см. рис. 7.9). Аналогично,

$$U_n(\delta,x_1^0,\ldots,x_n^0) \overset{(def)}{=} \{(x_1,x_2,\ldots,x_n)\,|\,(x_1-x_1^0)^2+\ldots+(x_n-x_n^0)^2<\delta^2\}.$$

Если при фиксированном числе $\delta>0$ точка $(x_1,x_2) \in U_2(\delta,x_1^0,x_2^0)$, то говорят, что точка (x_1,x_2) *близка* к точке (x_1^0, x_2^0). Если точка $(x_1,x_2) \notin U_2(\delta,x_1^0,x_2^0)$, то говорят, что точка (x_1,x_2) далека от точки (x_1^0,x_2^0). Если точка (x_1^0,x_2^0) принадлежит множеству M вместе со своей некоторой δ-окрестностью $U_2(\delta,x_1^0,x_2^1)$, т.е. со *всеми* своими близкими точками (x_1,x_2), она (точка (x_1^0,x_2^0) называется *внутренней* для множества M.

В случае, когда множество M есть неотрицательный ортант плоскости, т.е. $M=\{(x_1,x_2)|x_1 \geq 0,\ x_2 \geq 0\}$, то точки (x_1,x_2) с координатами $x_1>0$, $x_2>0$ - это внутренние точки неотрицательного ортанта. Точки (x_1,x_2) неотрицательного ортанта, у которых хотя бы одна координата равна нулю, являются граничными точками неотрицательного ортанта.

Определение 4.2.

Точка (x_1^0, x_2^0) называется *точкой локального максимума (минимума)* функции $f(x_1, x_2)$ двух переменных x_1 и x_2, если для всех точек (x_1, x_2) из области определения функции f, *близких* к точке (x_1^0, x_2^0), справедливо неравенство $f(x_1^0,x_2^0) \geq f(x_1,x_2)$ ($f(x_1^0,x_2^0) \leq f(x_1,x_2)$).

Само частное значение $f(x_1^0, x_2^0)$ называется *локальным максимумом (локальным минимумом)* функции $y = f(x_1, x_2)$.

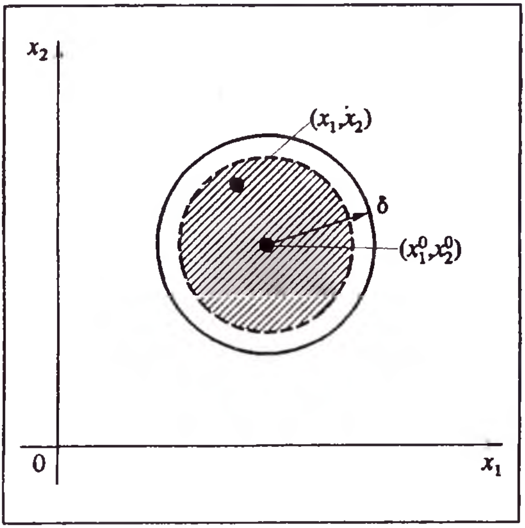


Рис. 7.9

Если (x_1^0, x_2^0) - точка локального максимума (минимума) функции $y = f(x_1, x_2)$, то около точки $(x_1^0, x_2^0, f(x_1^0, x_2^0))$ трехмерного пространства график $Г$ функции $y = f(x_1, x_2)$ имеет вид "шапочки" (перевернутой "шапочки") (см. рис. 7.10а и рис. 7.10б).

Отметим, что вместо двух терминов (максимума и минимума) используют один термин *экстремум*.

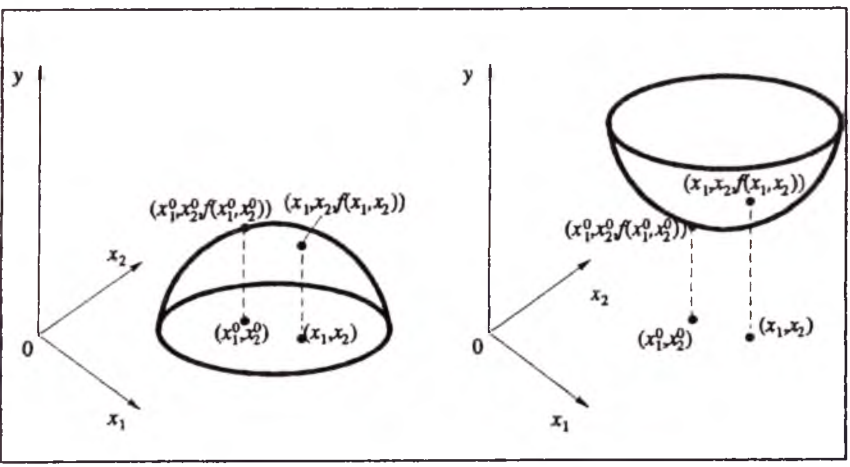


Рис. 7.10а *Рис. 7.10а*

Определение 4.3.

Точка (x_1^0, x_2^0) называется точкой *глобального максимума (глобального минимума)* функции $y=f(x_1,x_2)$ двух переменных x_1 и x_2, если для *всех* точек (x_1,x_2), для которых функция $f(x_1,x_2)$ определена, справедливо неравенство $f(x_1^0,x_2^0) \geq f(x_1,x_2)$ $(f(x_1^0,x_2^0) \leq f(x_1,x_2))$.

Само частное значение $f(x_1^0, x_2^0)$ называется *глобальным максимумом (глобальным минимумом)* функции $y = f(x_1, x_2)$.

Если функция $f(x_1,x_2)$ выпукла вниз и имеет локальный минимум, то он является глобальным минимумом. Если функция $f(x_1,x_2)$ выпукла вверх и имеет локальный максимум, то он является глобальным максимумом.

В экономической теории функция $f(x_1, x_2)$ обычно определена при $x_1 \geq 0$, $x_2 \geq 0$ и она либо выпукла вверх, либо выпукла вниз, поэтому её локальный максимум (локальный минимум) является также и глобальным.

Необходимое условие локального экстремума формулируется следующим образом.

Пусть функция $y = f(x_1, x_2)$ в точке (x_1^0, x_2^0) имеет локальный экстремум (точка (x_1^0, x_2^0) - внутренняя для области определения функции $y=f(x_1,x_2)$), тогда

$$\frac{\partial f(x_1^0, x_2^0)}{\partial x_1} = 0, \quad \frac{\partial f(x_1^0, x_2^0)}{\partial x_2} = 0$$

(существование (первых) частных производных в точке (x_1^0, x_2^0) предполагается).

Определение 4.4.

Точка (x_1^0, x_2^0) называется *критической* для функции $y = f(x_1, x_2)$, если координаты x_1^0 и x_2^0 этой точки удовлетворяют системе уравнений $\dfrac{\partial f(x_1, x_2)}{\partial x_1} = 0$, $\dfrac{\partial f(x_1, x_2)}{\partial x_2} = 0$.

Поэтому точки локального экстремума функции $y = f(x_1, x_2)$, лежащие внутри её области определения, следует искать только среди критических точек этой функции.

Критическая точка не обязана быть точкой (локального) экстремума, как показывает следующий пример.

Пример 4.1. Для функции $y=x_1^2-x_2^2$ имеем $\dfrac{\partial y}{\partial x_1} = 2x_1 = 0$, $\dfrac{\partial y}{\partial x_2} = -2x_2 = 0$, откуда получаем критическую точку (0,0) ($y(0,0)=0$). Однако точка (0, 0) не есть ни точка максимума, ни минимума, ибо при $(x_1,0)$ $y = x_1^2 > 0$, при $(0, x_2)$ $y = -x_2^2 < 0$, а при (0, 0) $y = 0$. График функции $y = x_1^2 - x_2^2$ называется *седловой поверхностью* (см. рис. 7.11, на котором хорошо видно, что около трехмерной точки (0, 0, 0) поверхность сильно отличается по своему виду от "шапочки" и перевернутой "шапочки").

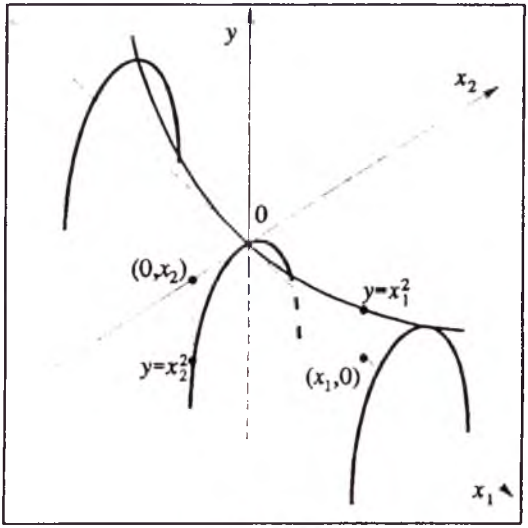


Рис. 7.11

Определение 4.5.

Второй частной производной функции $y = f(x_1, x_2)$ двух переменных называется (первая) частная производная от (первой) частной производной.

Таким образом имеем четыре вторых частных производных

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2},\ \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2},\ \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_1},\ \frac{\partial^2 f(x_1,x_2)}{\partial x_2^2}.$$

Если *смешанные* вторые частные производные $\frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2}$ и $\frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_1}$ непрерывны, то они обязательно равны. В отличие от смешанных вторые частные производные $\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2}$, $\frac{\partial^2 f(x_1,x_2)}{\partial x_2^2}$ принято называть *чистыми*.

В случае функции $f(x_1, \ldots, x_n)$ n переменных имеем n^2 штук вторых частных производных: $\frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_1^2}$, …, $\frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_n^2}$,

$$\frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_1 \partial x_2},\ \ldots,\ \frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_1 \partial x_n},\ \ldots,$$

$$\frac{\partial^2 f(x_1,\ldots x_n)}{\partial x_{n-1} \partial x_n},\ \frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_n \partial x_{n-1}}.$$

Если смешанные вторые частные производные $\frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_i \partial x_j}$ и $\frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_j \partial x_i}$ $(i \neq j;\ i,j=1,\ldots,n)$ непрерывны, то они равны.

Пример 4.2. Пусть функция $f(x_1,x_2)$ есть квадратичная форма $y=x_1^2-4x_1x_2+x_2^2$. Здесь $-\infty < x_i < +\infty$ $(i=1,2)$. Тогда $\frac{\partial y}{\partial x_1} = 2x_1 - 4x_2$, $\frac{\partial y}{\partial x_2} = -4x_1 + 2x_2$.

$$\frac{\partial^2 y}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial y}{\partial x_1}\right) = \frac{\partial(2x_1 - 4x_2)}{\partial x_1} = 2,$$

$$\frac{\partial^2 y}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial y}{\partial x_2}\right) = \frac{\partial(-4x_1 + 2x_2)}{\partial x_1} = -4,$$

$$\frac{\partial^2 y}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_2}\left(\frac{\partial y}{\partial x_1}\right) = \frac{\partial(2x_1 - 4x_2)}{\partial x_2} = -4,$$

$$\frac{\partial^2 y}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial y}{\partial x_2}\right) = \frac{\partial(-4x_1 + 2x_2)}{\partial x_2} = 2.$$

Достаточное условие локального экстремума формулируется следующим образом.

Пусть функция $y=f(x_1,x_2)$ имеет критическую точку (x_1^0,x_2^0) (т.е. пусть $\frac{\partial f(x_1^0,x_2^0)}{\partial x_1} = \frac{\partial f(x_1^0,x_2^0)}{\partial x_2} = 0$).

1) Пусть $\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1^2} > 0$ (или $\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_2^2} > 0$),

$\left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1^2}\right)\left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_2^2}\right) - \left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1 \partial x_2}\right)^2 > 0$, тогда (x_1^0,x_2^0) - точка локального минимума функции $y=f(x_1,x_2)$.

2) Пусть $\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1^2} < 0$ (или $\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_2^2} < 0$),

$$\left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1^2}\right)\left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_2^2}\right) - \left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1 \partial x_2}\right)^2 > 0,$$ тогда (x_1^0, x_2^0) - точка локального максимума функции $y = f(x_1, x_2)$,

3) Пусть $$\left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1^2}\right)\left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_2^2}\right) - \left(\frac{\partial^2 f(x_1^0,x_2^0)}{\partial x_1 \partial x_2}\right)^2 < 0,$$ тогда в точке (x_1^0,x_2^0) у функции $y=f(x_1,x_2)$ локального и, следовательно, глобального экстремума нет.

В приведённом достаточном условии предполагается, что точка (x_1^0, x_2^0) - внутренняя для области определения функции $f(x_1,x_2)$ и что вторые частные производные функции $f(x_1,x_2)$ определены в точке (x_1^0, x_2^0) и во всех близких к ней точках (x_1,x_2) и непрерывны в точке (x_1^0, x_2^0).

Пример 4.3. Продолжим пример 4.2. Имеем $\frac{\partial y}{\partial x_1} = 2x_1 - 4x_2 = 0$, $\frac{\partial y}{\partial x_2} = -4x_1 + 2x_2 = 0$, откуда получаем единственную критическую точку (0,0). Для этой точки (и любой другой точки (x_1,x_2)) имеем

$$\left(\frac{\partial^2 y}{\partial x_1^2}\right)\left(\frac{\partial^2 y}{\partial x_2^2}\right) - \left(\frac{\partial^2 y}{\partial x_1 \partial x_2}\right)^2 = -12 < 0,$$ т.е. в точке (0,0) локального и глобального экстремума нет.

Задача 4.1. Исследовать на экстремум следующую квадратичную функцию двух переменных: $y=x_1^2-2x_1x_2+2x_2^2-x_1-2x_2$. Эту задачу предлагается решить самостоятельно.

Определения локального и глобального экстремума и необходимое условие локального экстремума функции $y = f(x_1, \ldots, x_n)$ n переменных $x_1, \ldots, x_n$ повторяются почти дословно. В частности, необходимое условие локального экстремума имеет вид $\frac{\partial f(x_1,\ldots,x_n)}{\partial x_1} = 0, \ldots, \frac{\partial f(x_1,\ldots,x_n)}{\partial x_n} = 0$. Однако обобщение достаточного условия локального экстремума для случая функции n переменных является сложным и поэтому здесь не приводится, а рекомендуется лишь для дополнительного чтения (см. Интрилигатор. М., с. 71-74).

В заключение этого раздела отметим, что более точными являются термины: *безусловный* локальный максимум (минимум), точка *безусловного* локального максимума (минимума), *безусловный* глобальный максимум (минимум), точка *безусловного* глобального максимума (минимума). Вместо термина *безусловный* используется менее удачный (но достаточно выразительный) термин *абсолютный*.

Пример 4.4. Прибыль $PR(x_1,x_2)$ вычисляется по следующей формуле $PR(x_1,x_2)=p_0f(x_1,x_2)-p_1x_1-p_2x_2$, где $f(x_1,x_2)$ - производственная функция фирмы, p_0 - рыночная цена продукции, выпускаемой фирмой, p_1 и p_2 - соответственно рыночные цены первого и второго ресурсов (факторов производства). Выражение $p_0f(x_1,x_2)$ называется выручкой фирмы, $p_1x_1+p_2x_2$ - издержками производства фирмы, если для выпуска продукции фирма затрачивает первый и второй ресурсы в количествах x_1 и x_2 единиц. Задача ставится так. Определить комбинацию (x_1^0,x_2^0) ресурсов, при которой фирма получит наибольшую прибыль. Для решения этой задачи следует найти критические точки функции $PR(x_1,x_2)$, т.е. следует решить систему уравнений

$$\frac{\partial PR(x_1,x_2)}{\partial x_1} = p_0\frac{\partial f(x_1,x_2)}{\partial x_1} - p_1 = 0, \qquad \frac{\partial PR(x_1,x_2)}{\partial x_2} = p_0\frac{\partial f(x_1,x_2)}{\partial x_2} - p_2 = 0.$$

Поскольку производственная функция $y=f(x_1,x_2)$ обладает рядом специфических условий (в частности, если ее график напоминает горку - см. раздел 1), постольку часто критическая точка (x_1^0, x_2^0) является единственной и обязательно точкой (глобального) максимума прибыли $y = PR(x_1, x_2)$.

Вопросы к главе 7

1. Что называется графиком функции двух переменных? Приведите примеры (не менее двух) графиков.
2. Сформулируйте определение множества (линии) уровня функции двух переменных. Может ли множество уровня функции двух переменных не быть линией? Если может, приведите примеры. Могут ли множества (линии) двух различных уровней иметь общие точки? Дайте обоснование ответа.
3. Как в терминах линий уровня описать подъем туриста на гору (холм)?
4. В чем отличие (первой) частной производной от (первой) производной?
5. Опишите взаимосвязь между градиентом функции двух переменных и ее линией уровня.
6. Приведите формулу производной неявной функции.
7. Сформулируйте определение (первого) полного дифференциала.
8. Сформулируйте определение однородной функции степени p.
9. В чем суть теоремы Эйлера?
10. Сформулируйте определение локального и глобального экстремума функции двух и n переменных. Может ли глобальный экстремум не быть локальным?

ГЛАВА 8
ОПТИМИЗАЦИОННЫЕ ЗАДАЧИ С ОГРАНИЧЕНИЯМИ

8.1. Задачи на условный экстремум

В теории безусловного локального экстремума сравнивают частное значение $f(x_1^0,x_2^0)$ функции $y = f(x_1,x_2)$ в точке (x_1^0,x_2^0) с частными значениями $f(x_1,x_2)$ этой функции во *всех* точках (x_1,x_2), *близких* к точке (x_1^0,x_2^0) (см. главу 7, раздел 7.4). Другими словами, в теории локального безусловного экстремума на независимые переменные x_1 и x_2 не накладываются никакие *дополнительные условия*, т.е. не требуется, чтобы переменные x_1 и x_2 удовлетворяли некоторым *дополнительным ограничениям*.

Рассмотрим теперь *другую* задачу.

Найти локальный максимум (или локальный минимум) функции $y = f(x_1,x_2)$ при *условии*, что независимые переменные x_1 и x_2 удовлетворяют *ограничению* $g(x_1,x_2) = 0$ в виде равенства, т.е.

$$f(x_1,x_2) \to \max \qquad (f(x_1,x_2) \to \min) \tag{1}$$

при условии

$$g(x_1,x_2) = 0. \tag{2}$$

Задача (1), (2) называется задачей на *условный* локальный максимум (минимум). Термин *условный* здесь появляется в связи с тем, что независимые переменные x_1 и x_2 удовлетворяют *условию* (ограничению) (2). Вместо двух терминов (максимум и минимум) используется обобщенный термин *экстремум*. В задаче (1), (2) на условный экстремум функцию $f(x_1,x_2)$ принято называть *целевой*, ибо ее максимизация (или минимизация) часто есть формальное выражение какой-то *цели* (например, максимизации объема производства при фиксированных затратах). Функцию g называют функцией, задающей ограничение, или функцией связи.

Уравнение (2) есть уравнение *нулевой* линии (точнее множества) уровня функции $g(x_1,x_2)$, ибо $g(x_1,x_2)=\tau$, где $\tau=0$. Поэтому задачу на условный локальный максимум (минимум) можно еще сформулировать так: среди точек нулевой линии уровня функции $y=g(x_1,x_2)$ найти точку (x_1^0,x_2^0), в которой частное значение $f(x_1^0,x_2^0)$ функции $y=f(x_1,x_2)$ больше (или меньше) ее частных значений $f(x_1,x_2)$ в остальных точках (x_1,x_2) этой линии, *близких* к точке (x_1^0,x_2^0) (см. рис. 8.1). Точка (x_1^0,x_2^0) называется точкой *условного локального максимума (минимума)* функции $f(x_1,x_2)$, само частное значение $f(x_1^0,x_2^0)$ - *условным локальным максимумом (минимумом)* функции $f(x_1,x_2)$ при наличии ограничения $g(x_1,x_2)=0$.

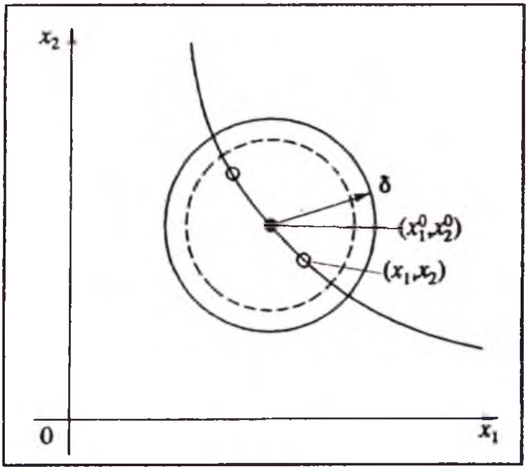


Рис. 8.1

Проиллюстрируем задачу на условный максимум в трехмерном пространстве (см. рис. 8.2).

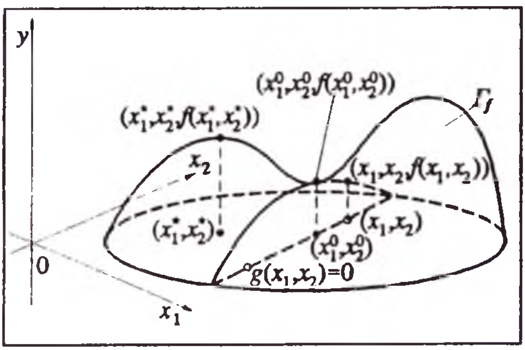


Рис. 8.2

Точка (x_1^*,x_2^*) - точка абсолютного локального максимума функции $f(x_1,x_2)$, ибо точка $(x_1^*,x_2^*,f(x_1^*,x_2^*))$ - локальная (т.е. местная) “макушка” графика Γ_f этой функции $f(x_1,x_2)$. Точка (x_1^0, x_2^0) - точка условного локального максимума функции $f(x_1,x_2)$, ибо точка $(x_1^0,$

$x_2^0, f(x_1^0,x_2^0))$ - самая высокая точка "тропинки" L, которая проходит через "перевал" графика Γ_f. На рис. 8.2 четко видно, что точка $(x_1^0,x_2^0,f(x_1^0,x_2^0))$ "макушкой" не является, т.е. точка (x_1^0,x_2^0) условного локального максимума может не быть точкой безусловного локального максимума. Линия $g(x_1,x_2)=0$ есть проекция "тропинки" L на координатную плоскость Ox_1x_2.

Отметим, что если известен график Γ_f функции $f(x_1,x_2)$ двух переменных x_1 и x_2, то, глядя на него, можно сразу понять, есть ли точки абсолютного и условного локального экстремума или какие-то из них (а, возможно, все) отсутствуют.

В случае, если графика Γ_f нет, точки абсолютного локального экстремума отыскиваются с помощью формального анализа функции $y = f(x_1,x_2)$, который был описан в главе 7.

Существует формальный метод отыскания точек условного локального экстремума, для использования которого также не требуется знания графика Γ_f функции $y = f(x_1,x_2)$ и графика уравнения $g(x_1,x_2)=0$. Этот метод (он называется методом Лагранжа) подробно описан в следующем разделе 2.

Если значение $f(x_1^0,x_2^0)$ функции $f(x_1,x_2)$ больше (меньше) значений $f(x_1,x_2)$ этой функции во *всех* точках (x_1,x_2) линии $g(x_1,x_2)=0$, то значение $f(x_1^0,x_2^0)$ называется условным *глобальным* максимумом (минимумом) функции $f(x_1,x_2)$ при наличии ограничения $g(x_1,x_2)=0$, а точка (x_1^0,x_2^0) - точкой условного *глобального* максимума (минимума) функции $f(x_1,x_2)$.

Точка условного глобального максимума (минимума) функции $f(x_1,x_2)$ является точкой условного *локального* максимума (минимума) этой функции. Обратное, вообще говоря, неверно. На рис. 8.2 точка (x_1^0,x_2^0) является точкой не только локального, но и глобального условного максимума функции $f(x_1,x_2)$ при наличии ограничения $g(x_1,x_2) = 0$.

В случае функции $f(x_1, ...,x_n)$ n независимых переменных $x_1, ..., x_n$ задача на условный максимум (минимум) формулируется так:

$$f(x_1, ..., x_n) \max \quad (f(x_1, ..., x_n) \to \min) \tag{3}$$

при условиях

$$\begin{matrix} g_1(x_1, ..., x_n) = 0 \\ \dots\dots\dots\dots \\ g_m(x_1, ..., x_n) = 0 \end{matrix} \tag{4}$$

(обычно $m<n$).

Если частное значение $f(x_1^0, ..., x_n^0)$ сравниваются со значениями $f(x_1, ..., x_n)$ в точках $(x_1, ..., x_n)$, удовлетворяющих уравнениям (4) и *близких* к точке $(x_1^0, ..., x_n^0)$, то имеем задачу на условный *локальный* экстремум (максимум или минимум) функции $f(x_1, ..., x_n)$.

Если значение $f(x_1^0, ..., x_n^0)$ сравнивается с значениями во всех точках $(x_1, ..., x_n)$, удовлетворяющих уравнениям (4), то имеем задачу на условный *глобальный* экстремум (максимум или минимум) функции $f(x_1, ..., x_n)$.

Теория условного экстремума активно используется в микро- и макроэкономической теории. В задачах этой теории обычно локальный условный экстремум является также и глобальным условным экстремумом. Разберем конкретный пример.

Пример 1.1. Найти экстремум функции

$$y = x_1^2 + x_2^2 \tag{5}$$

при условии, что

$$x_1 + x_2 - 1 = 0, \tag{6}$$

т.е. решить задачу на условный экстремум.

Решение примера 1.1. Отметим прежде всего, что экстремум (экстремумы) функции (5) отыскиваются не на всей плоскости Ox_1x_2 (как это было в главе 7), а только на прямой (6).

Естественным является следующий способ решения задачи (5),(6) на условный экстремум. Выразить из уравнения (6) переменную x_2 через переменную x_1 и подставить полученное выражение $x_2 = 1 - x_1$ в функцию (5). Тогда задача (5),(6) на условный экстремум функции (5) *двух* переменных сведется к задаче на безусловный экстремум функции $y = 2x_1^2 - 2x_1 + 1$ *одной* переменной x_1.

Для решения задачи на безусловный экстремум найдем первую производную $y'=4x_1-2$ функции $y=2x_1^2-2x_1+1$ и приравняем первую производную к нулю: $4x_1-2=0$, откуда получим, что $x_1^0=½$. При переходе (слева направо) переменной x_1 через точку x_1^0 первая производная y' меняет знак с минуса на плюс, поэтому критическая точка x_1^0 есть точка локального минимума функции $y=2x_1^2-2x_1+1$. Очевидно, этот локальный минимум $y^0=2(x_1^0)^2-2x_1^0+1=½$ является также глобальным (см. на рис. 8.3 линию H, которая есть график функции $y=2x_1^2-2x_1+1$). Других локальных и глобальных экстремумов функция $y=2x_1^2-2x_1+1$ не имеет, ибо не существует точек, отличных от точки x_1^0, в которых бы производная $y=4x_1-2$ обращалась в нуль.

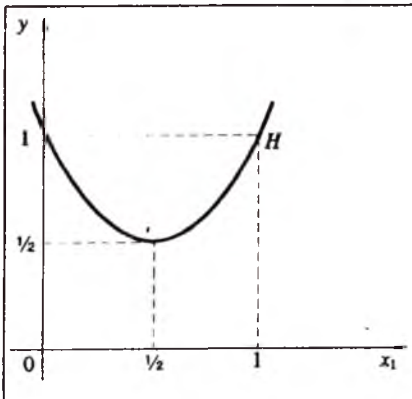


Рис. 8.3

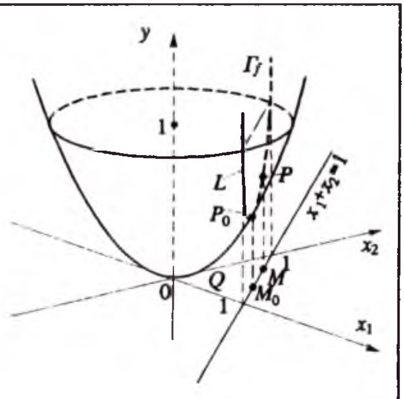


Рис. 8.4

Из полученного следует, что $(x_1^0,x_2^0)=(½,½)$ - точка условного глобального минимума функции (5), сам условный минимум равен $y^0=(x_1^0)^2+(x_1^0)^2=½$.

На рис. 8.4 дана геометрическая иллюстрация решения задачи (5),(6). На линии L, по которой пересекаются вертикальная плоскость Q и график Γ_f функции (5), самой низкой точкой является точка $P_0=(x_1^0,x_2^0,y^0)=(½,½,½)$. На поверхности Γ_f самой низкой является точка $0 = (0,0,0)$. Таким образом, на рис. 4 видно, что условный глобальный минимум функции (5), который равен ½ не совпадает с ее абсолютным (безусловным) минимумом, равным нулю. На рис. 4 также хорошо видно, что ни на линии L, ни на графике Γ_f нет самых высоких точек, т.е. функция (5) не имеет условного глобального максимума и абсолютного глобального максимума.

Решение примера 1.1 подсказывает следующий естественный на первый взгляд способ решения задачи (1),(2). С помощью уравнения (2) сначала выразить переменную x_2 через переменную x_1 (или переменную x_1 через переменную x_2). Затем полученное выражение $x_2 = h(x_1)$ подставить в функцию (1), которая после этого станет функцией $f(x_1,h(x_1))$ одной переменной x_1, и эту функцию исследовать на (безусловный) экстремум. Из отсутствия точки (точек) экстремума у функции $f(x_1,h(x_1))$ следует отсутствие точки (точек) условного экстремума у функции (1). Если x_1^0 - точка экстремума функции $y = f(x_1, h(x_1))$, то точка $(x_1^0, x_2^0) = (x_1^0, h(x_1^0))$ - точка условного экстремума функции (1) при наличии ограничения (2).

Однако, к сожалению, выразить аналитически переменную x_2 через переменную x_1 (или переменную x_1 через переменную x_2) часто бывает сложно, а то и невозможно. По этой причине только что описанная простая идея сведения задачи на условный экстремум для функции (1) *двух* переменных к задаче на безусловный экстремум для функции $f(x_1, h(x_1))$ *одной* переменной не может быть использована в качестве основы универсального метода решения задачи (1),(2) на условный экстремум.

В конце XVIII века Лагранж предложил остроумный метод решения задачи (1), (2) на условный экстремум, в котором не следует прибегать к разрешению уравнения (2) относительно одной переменной при фиксированной другой переменной. В этом методе число независимых переменных не сокращается, а, наоборот, растет.

Метод Лагранжа позволяет решать не только задачи вида (1),(2), но более общие задачи вида (3),(4).

Метод Лагранжа описан в следующем разделе.

8.2. Метод Лагранжа решения задачи на условный экстремум

Суть метода Лагранжа состоит в построении функции

$$L(x_1,x_2,\lambda) = f(x_1,x_2) + \lambda g(x_1,x_2) \tag{7}$$

трех переменных x_1,x_2,λ (называемой *функцией Лагранжа)* и, грубо говоря, в *сведении* задачи (1),(2) на условный экстремум в случае двух независимых переменных к задаче на абсолютный экстремум функции $L(x_1,x_2,\lambda)$ трех независимых переменных x_1, x_2, λ. Ниже эта идея о сведении уточняется.

Функция Лагранжа $L(x_1,x_2,\lambda)$ представляет собой сумму *целевой функции* (1) и *функции ограничения* (2), умноженной на новую независимую переменную λ (называемую *множителем Лагранжа)*, входящую обязательно в первой степени.

Необходимое условие локального условного экстремума функции (1) при наличии ограничения (2) *в аналитической форме* имеет вид.

Пусть функции $f(x_1,x_2)$, $g(x_1,x_2)$ непрерывны и имеют непрерывные частные производные первого порядка по переменным x_1 и x_2; пусть (x_1^0,x_2^0) - точка условного локального экстремума функции (1) при наличии ограничения (2) и пусть $\operatorname{grad} g(x_1^0,x_2^0) \neq 0$.

Тогда существует единственное число λ^0 такое, что (трехмерная) точка (x_1^0,x_2^0,λ^0) удовлетворяет следующей системе трех уравнений

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial x_1}=0,\quad \frac{\partial L(x_1,x_2,\lambda)}{\partial x_2}=0,\quad \frac{\partial L(x_1,x_2,\lambda)}{\partial \lambda}=0 \tag{8}$$

с тремя неизвестными x_1, x_2, λ (отметим, что всегда $\frac{\partial L(x_1,x_2,\lambda)}{\partial \lambda} = g(x_1,x_2)$).

Другими словами, если (двумерная) точка (x_1^0,x_2^0) есть точка локального условного экстремума функции (1) при наличии ограничения (2), то (трехмерная) точка (x_1^0,x_2^0,λ^0) - критическая точка функции Лагранжа. Отсюда следует, что для нахождения точек (условного) локального экстремума функции (1) при наличии ограничения (2) следует прежде всего найти критические точки функции Лагранжа, т.е. найти все решения системы уравнений (8). Далее критические точки функции Лагранжа следует "укоротить", удалив из них последние координаты λ. Затем каждую "укороченную" критическую точку необходимо проанализировать на предмет, является ли она в действительности точкой (условного) локального экстремума функции (1) при наличии ограничения (2) или не является.

Достаточное условие локального условного экстремума функции (1) при наличии ограничения (2) здесь не приводится. При анализе "укороченной" критической точки обычно используют наглядные геометрические или содержательные (экономические) соображения.

Отметим, что в некоторых задачах на условный экстремум, которые появляются в экономической теории, обычно "укороченная" критическая точка функции Лагранжа является на самом деле точкой условного локального (в действительности и глобального) экстремума функции (1) при наличии ограничения (2).

Пример 1.1. (продолжение). Решим задачу (5),(6) на условный экстремум методом Лагранжа. Имеем

$$L(x_1,x_2,\lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1),$$

откуда

$$\frac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0,\ \frac{\partial L}{\partial x_2} = 2x_2 + \lambda = 0,\ \frac{\partial L}{\partial \lambda} = x_1 + x_2 - 1 = 0. \quad (9)$$

Из первых двух уравнений вытекает, что $-2x_1 = \lambda = -2x_2$, т.е. $x_1 = x_2$, откуда, используя третье уравнение, получаем, что $x_1^0=x_2^0=½$.

Таким образом, система уравнений (9) имеет единственное решение, т.е. дает единственную критическую точку функции Лагранжа (½,½,-1) ($\lambda^0=-2x_1^0=-2$ ½$=-1$). "Укороченная" критическая точка (x_1^0,x_2^0)=(½,½) есть точка условного локального (также глобального) минимума функции (5) при наличии ограничения (6), ибо непосредственно проверяется, что при (x_1,x_2) (x_1^0,x_2^0) и удовлетворяющей уравнению (6) справедливо неравенство $f(x_1,x_2)>f(x_1^0,x_2^0)$=½.

$$\frac{\partial L}{\partial x_1}=0,\ldots,\ \frac{\partial L}{\partial x_n}=0,\ \frac{\partial L}{\partial \lambda_1}=0,\ldots,\ \frac{\partial L}{\partial \lambda_m}=0. \quad (10)$$

В случае общей задачи (3),(4) на условный экстремум функция Лагранжа имеет вид $L(x_1, ..., x_n, \lambda_1, ..., \lambda_m) = f(x_1, ..., x_n) + \lambda_1 g_1(x_1, ..., x_n) + ... + \lambda_m g_m(x_1, ..., x_m)$, а система (8) переписывается в виде системы $n + m$ уравнений с $n + m$ неизвестными $x_1, ..., x_n, \lambda_1, ..., \lambda_m$.

Критическая $(n + m)$-мерная точка $(x_1^0, ..., x_n^0, \lambda_1^0, ..., \lambda_m^0)$ функции Лагранжа после операции "укорачивания" приобретает вид $(x_1^0, ..., x_n^0)$ n-мерной точки.

Возвращаемся к случаю двух переменных x_1 и x_2. Необходимое условие (в виде системы уравнений (8)) локального условного экстремума перепишем в развернутом виде:

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial x_1} = \frac{\partial f(x_1,x_2)}{\partial x_1} + \lambda \frac{\partial g(x_1,x_2)}{\partial x_1} = 0, \quad (11.1)$$

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial x_2} = \frac{\partial f(x_1,x_2)}{\partial x_2} + \lambda \frac{\partial g(x_1,x_2)}{\partial x_2} = 0, \quad (11.2)$$

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial \lambda}=g(x_1,x_2)=0. \qquad (11.3)$$

Поскольку grad $f(x_1,x_2) = \left(\frac{\partial f(x_1,x_2)}{\partial x_1}, \frac{\partial f(x_1,x_2)}{\partial x_2}\right)$, grad $g(x_1,x_2) = \left(\frac{\partial g(x_1,x_2)}{\partial x_1}, \frac{\partial g(x_1,x_2)}{\partial x_2}\right)$, уравнения (11.1) и (11.2) можно представить в компактной векторной форме

$$\text{grad} f(x_1,x_2) + \lambda \text{grad}\, g(x_1,x_2) = 0. \qquad (12)$$

Для критической точки (x_1^0,x_2^0,λ^0) функции Лагранжа имеем

$$\text{grad} f(x_1^0,x_2^0) + \lambda^0 \text{grad}(x_1^0,x_2^0) = 0, \qquad (13)$$

т.е. в "укороченной" критической точке (x_1^0,x_2^0) функции Лагранжа

$$\text{grad} f(x_1^0,x_2^0) = - \lambda^0 \text{grad}(x_1^0,x_2^0), \qquad (14)$$

что эквивалентно тому, что в точке (x_1^0,x_2^0) линии уровней $f(x_1^0,x_2^0)$ и 0 функций $f(x_1,x_2)$ и $g(x_1,x_2)$ соответственно *касаются* (напомним, что grad$(x_1^0,x_2^0)\neq 0$).

Теперь приведем необходимое условие локального условного экстремума функции (1) при наличии ограничения (2) в *геометрической форме*.

Пусть функции $f(x_1,x_2)$, $g(x_1,x_2)$ непрерывны и имеют непрерывные частные производные первого порядка по переменным x_1 и x_2; пусть (x_1^0,x_2^0) - точка условного локального экстремума функции (1) при наличии ограничения (2) и пусть

$$\text{grad} f(x_1^0,x_2^0) \neq 0,\ \text{grad}\, g(x_1^0,x_2^0) \neq 0.$$

Тогда градиенты grad $f(x_1^0,x_2^0)$ и grad $g(x_1^0,x_2^0)$, выходящие из точки (x_1^0,x_2^0), обязательно расположены на одной прямой, что эквивалентно тому, что *линии уровней* функций $f(x_1,x_2)$ и $g(x_1,x_2)$, содержащие точку (x_1^0,x_2^0), *касаются* в этой точке (см. рис.8.5).

Точка (x_1^0,x_2^0) на рис. 8.5 - точка условного локального максимума, что очевидно на основании наглядных геометрических соображений, ибо $\tau_1<\tau_0<\tau_2$ (grad $f(x_1^0,x_2^0)$ указывает направление наибыстрейшего возрастания функции в точке (x_1^0,x_2^0)), поэтому

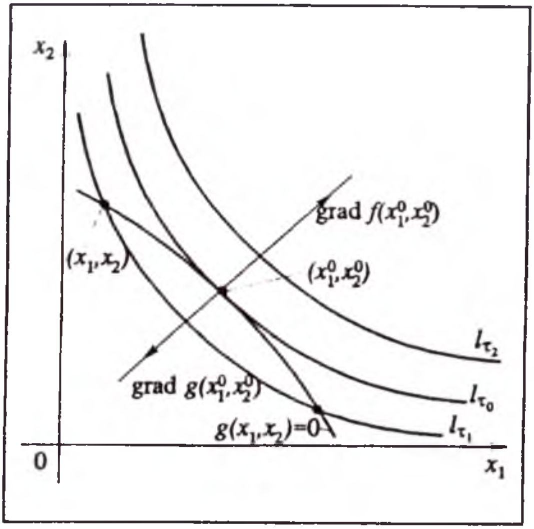


Рис. 8.5

$f(x_1^0,x_2^0)=\tau>\tau_1=f(x_1,x_2)$, если точка (x_1,x_2) принадлежит нулевому множеству уровня функции $g(x_1,x_2)$ и не совпадает с точкой (x_1^0,x_2^0). В случае, представленном на рис.5, имеем $\lambda^0\approx 2$. Рис. 8.5 близок к ситуации, типичной для экономической теории: градиент grad $f(x_1^0,x_2^0)$ целевой функции $f(x_1,x_2)$ "смотрит" на северо-восток, градиент grad $g(x_1^0,x_2^0)$ ограничения $g(x_1,x_2)$ "смотрит" либо на юго-запад, либо на северо-восток; фрагмент карты линий уровня целевой функции $f(x_1,x_2)$ похож на фрагмент типичной карты линий уровня, который встречается в экономической теории.

Необходимое условие (в том числе и геометрическое) локального экстремума функции (1) при наличии ограничения (2), вообще говоря, не является достаточным, т.е. в случае касания в точке (x_1^0,x_2^0) линий уровня функций $f(x_1,x_2)$ и $g(x_1,x_2)$ (что эквивалентно расположению на одной прямой градиентов grad $f(x_1^0,x_2^0)$ и grad $g(x_1^0,x_2^0)$, исходящих из точки (x_1^0,x_2^0)), точка (x_1^0,x_2^0) может и не быть точкой условного локального экстремума функции (1) при наличии ограничения (2) (см. рис. 8.6).

Точка (x_1^0,x_2^0) на рис. 8.6 -"укороченная" критическая точка функции Лагранжа, которая не является точкой локального условного экстремума функции (1) при наличии ограничения (2), что очевидно на основании наглядных геометрических соображений, ибо в точках (x_1,x_2), расположенных на линии $g(x_1,x_2)=0$ строго выше точки (x_1^0,x_2^0), справедливо неравенство $f(x_1,x_2)<f(x_1^0,x_2^0)$ $(\tau_1<\tau_0)$, а в точках (x_1,x_2), расположенных на линии $g(x_1,x_2)=0$ строго ниже точки (x_1^0,x_2^0), справедливо неравенство $f(x_1,x_2)>f(x_1^0,x_2^0)$ $(\tau_2>\tau_0)$. Для рис. 8.6 $\lambda^0\approx -½$.

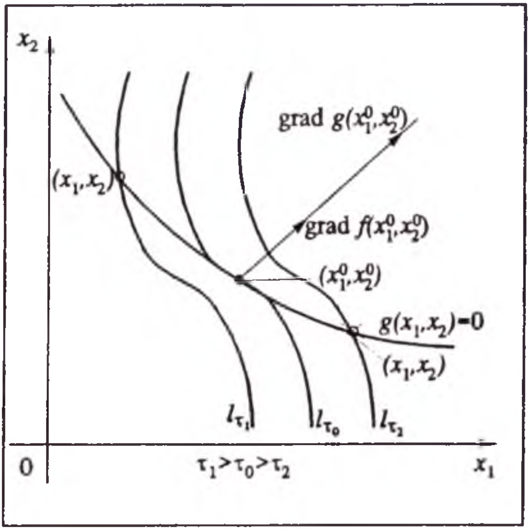


Рис. 8.6

Отметим, что фрагмент карты линий уровня целевой функции $f(x_1,x_2)$, представленный на рис. 8.6, не типичен для экономической теории.

Пример 1.1 (продолжение). Приведем аналог рис. 8.5 для задачи (5),(6) на условный экстремум (см. рис. 8.7).

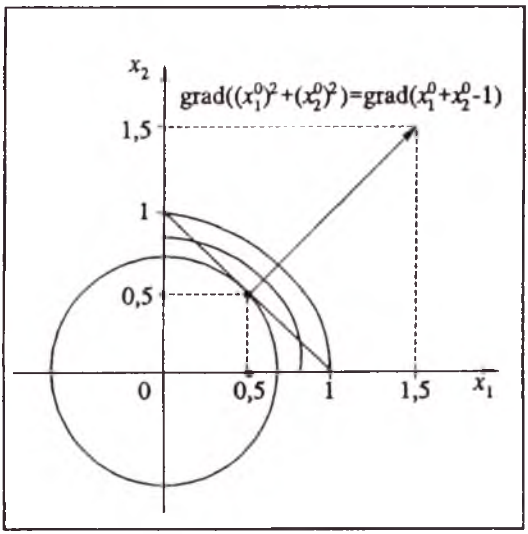


Рис. 8.7

В случае рис. 8.7:

$\mathrm{grad}(x_1^0))^2+(x_2^0)^2)=(2x_1^0,2x_2^0)=(2\cdot½,2\cdot½)=(1,1)$, $\mathrm{grad}(x_1^0+x_2^0-1)=(1,1)$, $\lambda^0=-1$.

8.3. Понятие о задаче математического программирования

Если в задаче (1),(2) на условный экстремум ограничение (2) в виде *равенства* заменяется на ограничение $g(x_1,x_2) \le 0$ в виде *неравенства,* то мы получаем частный случай задачи математического программирования:

$$f(x_1,x_2) \to \max \quad (f(x_1,x_2) \to \min) \tag{1}$$

при условии

$$g(x_1,x_2) \le 0. \tag{15}$$

В случае функции двух переменных задача математического программирования (для определенности - задача на максимум) имеет вид ((1),(16.1)-(16.m),(17))

$$f(x_1,x_2) \to \max \tag{1}$$

при условиях

$$g_1(x_1,x_2) \le 0 \qquad (16.1)$$

$$\dots\dots\dots\dots \quad \dots\dots$$

$$g_m(x_1,x_2) \le 0 \qquad (16.m)$$

$$x_1 \ge 0,\ x_2 \ge 0 \tag{17}$$

Функцию $f(x_1,x_2)$ принято называть *целевой,* неравенства (16.1)-(16.m) - *специальными ограничениями* задачи математического программирования, неравенства $x_1 \ge 0$, $x_2 \ge 0$ - *общими ограничениями* задачи математического программирования.

Точка (x_1,x_2), удовлетворяющая специальным и общим ограничениям, называется *допустимым решением* задачи математического программирования.

Множество всех допустимых решений задачи математического программирования (далее, для краткости, ЗМП) называется *допустимым множеством* этой задачи.

Если ЗМП имеет хотя бы одно допустимое решение (т.е. ее допустимое множество *не пусто),* она называется *допустимой,* если ЗМП не имеет ни одного допустимого решения (т.е. ее допустимое множество *пусто)*, она называется *недопустимой.*

Точка (x_1^0,x_2^0) называется *оптимальным решением* ЗМП, если, во-первых, она есть допустимое решение этой ЗМП и если, во-вторых, на этой точке целевая функция достигает *глобального максимума* (в случае задачи максимизации) или *глобального минимума* (в случае задачи минимизации) среди всех точек, удовлетворяющих ограничениям, т.е. для всех (x_1,x_2), удовлетворяющих неравенствам (16.1)-(16.m), (17), справедливо неравенство

$f(x_1^0,x_2^0) \geq f(x_1,x_2)$ (в случае задачи максимизации),
$f(x_1^0,x_2^0) \leq f(x_1,x_2)$ (в случае задачи минимизации).

На первый взгляд, ЗМП может рассматриваться как задача более общая по сравнению с задачами на абсолютный (если убрать все специальные и общие ограничения) и условный (если убрать все общие ограничения, а из специальных оставить одно в виде равенства) экстремумы. Однако в действительности полное обобщение места не имеет, ибо в случае ЗМП речь идет только о *глобальном* экстремуме, в то время как в случае задачи на абсолютный и условный экстремум речь идет как о глобальном, так и о локальном экстремуме.

В экономической теории часто ЗМП сводится к задаче на условный экстремум. В качестве иллюстрации приведем пример задачи потребительского выбора (задачи рационального поведения потребителя на рынке) в виде ЗМП (задача потребительского выбора подробно описана в следующей главе)

$$u(x_1,x_2) \to \max \tag{18}$$

при условиях

$$p_1x_1 + p_2x_2 \leq I \tag{19}$$
$$x_1 \geq 0,\ x_2 \geq 0. \tag{20}$$

Здесь (x_1,x_2) - потребительский набор (x_1 - число единиц первого продукта, x_2 - число единиц второго), p_1 - рыночная цена одной единицы первого продукта, p_2 - рыночная цена одной единицы второго продукта, I - доход индивидуума, который он готов потратить для приобретения продуктов, $u(x_1,x_2)$ - функция полезности индивидуума. Отметим, что первые частные производные функции полезности $u(x_1,x_2)$ по переменным x_1 и x_2, т.е. и , называются предельными полезностями первого и второго продукта соответственно.

Дадим геометрическую интерпретацию задачи потребительского выбора (см. рис. 8.8).

Заштрихованный треугольник на рис. 8.8 показывает множество потребительских наборов (x_1,x_2), которые доступны индивидууму, однако только на единственном потребительском наборе (x_1^0,x_2^0) потребитель максимизирует свою функцию полезности $u(x_1,x_2)$. В точке (x_1^0,x_2^0) бюджетная прямая $p_1x_1+p_2x_2=I$ и линия безразличия *касаются*. В связи с тем, что $p_1x_1^0 + p_2x_2^0 = I$, оптимальное решение (x_1^0,x_2^0) ЗМП совпадает с решением (x_1^0,x_2^0) следующей (более простой) задачи на условный (глобальный) экстремум

$$u(x_1,x_2) \to \max \tag{18}$$

при условии

$$p_1x_1+p_2x_2-I=0. \tag{21}$$

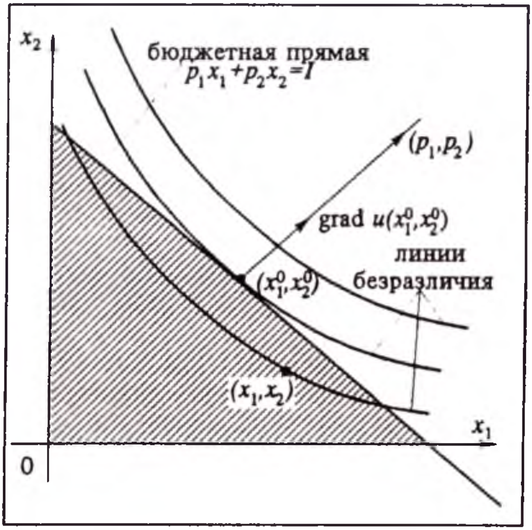


Рис. 8.8

Таким образом, задача потребительского выбора может быть описана как в виде ЗМП (18)-(20), так и в виде задачи на условный экстремум (18),(21). С математической точки зрения это *разные* задачи, однако они имеют *одно и то же решение* (x_1^0,x_2^0) - потребительский набор, который максимизирует (глобально) функцию полезности $u(x_1,x_2)$ и удовлетворяет бюджетному ограничению $p_1x_1+p_2x_2$ I как равенству: $p_1x_1^0+p_2x_2^0=I$. На рис. 8.8 также показаны градиенты функции полезности $u(x_1,x_2)$ и функции ограничения $p_1x_1+p_2x_2-I$ в точке (x_1^0,x_2^0): grad(x_1^0,x_2^0) и (p_1,p_2). Эти градиенты расположены на одной прямой, проходящей через точку (x_1^0,x_2^0), что, как уже отмечалось, эквивалентно *касанию* линии безразличия и бюджетной прямой в точке (x_1^0,x_2^0).

Из сказанного следует, что конкретные задачи рационального поведения потребителя на рынке можно решать (к сожалению, не всегда) как задачи на условный экстремум вида (18), (21).

Если в ЗМП все функции $f(x_1,x_2)$, $g(x_1,x_2)$, ..., $g_m(x_1,x_2)$ линейны, имеем *задачу линейного программирования* (ЗЛП), если хотя бы одна из функций $f(x_1,x_2)$, $g_1(x_1,x_2)$, ..., $g_m(x_1,x_2)$ нелинейная, имеем *задачу нелинейного программирования* (ЗНЛП).

ЗЛП на максимум в случае двух переменных x_1 и x_2 имеет вид

$$c_1x_1+c_2x_2 = v \to \max \qquad (22)$$

при условиях

$$a_{11}x_1+a_{12}x_2 \le b_1 \qquad (23.1)$$

....

$$a_{m1}x_1+a_{m2}x_2 \le b_m, \tag{23.m}$$
$$x_1 \ge 0,\ x_2 \ge 0 \tag{24}$$
(ЗЛП в стандартной форме)

или такой вид

$$c_1x_1+c_2x_2=v \to \max \tag{22}$$
при условиях
$$a_{11}x_1+a_{12}x_2=b_1 \tag{25}$$
$$x_1 \ge 0,\ x_2 \ge 0$$
(ЗЛП в канонической форме).

В ЗЛП числа $c_1,c_2,b_1,b_2,\ldots,b_m,a_{11},a_{12},\ldots,a_{m1},a_{m2}$ заданы.
В случае n переменных $x_1,\ldots,x_n$ ЗЛП на максимум имеет вид

$$c_1x_1+\ldots+c_nx_n=v \to \max \tag{26}$$
при условиях
$$a_{11}x_1+\ldots+a_{1n}x_n \le b_1 \tag{27.1}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots \quad \ldots\ldots$$
$$a_{m1}x_1+\ldots+a_{mn}x_n \le b_m \tag{27.m}$$
$$x_1 \ge 0,\ \ldots,\ x_n \ge 0 \tag{28}$$
(ЗЛП в стандартной форме на максимум)

или такой вид
$$c_1x_1+\ldots+c_nx_n=v \to \max \tag{26}$$
при условиях
$$a_{11}x_1+\ldots+a_{1n}x_n=b_1 \tag{29.1}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots \quad \ldots\ldots$$
$$a_{m1}x_1+\ldots+a_{mn}x_n=b_m \tag{29.m}$$
$$x_1 \ge 0,\ldots,\ x_n \ge 0 \tag{28}$$
(ЗЛП в канонической форме на максимум, здесь $m<n$).

ЗЛП на минимум имеет вид

$$c_1x_1+\ldots+c_nx_n=w \to \min$$
при условиях
$$a_{11}x_1+\ldots+a_{1n}x_n \ge b_1$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{m1}x_1+\ldots+a_{mn}x_n \ge b_m$$
$$x_1 \ge 0,\ \ldots,\ x_n \ge 0$$
(ЗЛП в стандартной форме на минимум)

или такой вид

$$c_1x_1+\ldots+c_nx_n=w \to \min$$

при условиях

$$a_{11}x_1+\ldots+a_{1n}x_n=b_1$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{m1}x_1+\ldots+a_{mn}x_n=b_m,$$
$$x_1 \geq 0,\ \ldots,\ x_n \geq 0$$

(ЗЛП в канонической форме на минимум, здесь $m<n$).

Следующая ЗЛП (в стандартной форме на минимум)

$$b_1p_1 + \ldots + b_mp_m = w \to \min \tag{30}$$

при условиях

$$a_{11}p_1 + \ldots + a_{m1}p_m \geq c_1 \tag{31.1}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots \quad \ldots\ldots$$
$$a_{1n}p_1 + \ldots + a_{mn}p_m \geq c_n, \tag{31.n}$$
$$p_1 \geq 0,\ \ldots,\ p_m \geq 0 \tag{32}$$

называется *сопряженной* (или *двойственной)* к ЗЛП (26), (27.1) - (27.m), (28), которая называется *исходной*. Обе ЗЛП (исходная (26), (27.1) - (27.m), (28) и сопряженная (30), (31.1) - (31.n), (32)) образуют *двойственную пару* ЗЛП. Сопряженная ЗЛП к сопряженной ЗЛП (30), (31.1) - (31.n), (32) совпадает с исходной ЗЛП (26), (27.1) - (27.m), (28). ЗЛП, сопряженная к ЗЛП (26), (29.1) - (29.m), (28) в канонической форме, имеет вид (30), (31.1) - (31.n). Здесь общие ограничения (32) не используются. Сопряженная к ЗЛП (30), (31.1) - (31.n) совпадает с исходной ЗЛП (26), (29.1) - (29.m), (28).

Вопросы к главе 8

1. Что такое задача на условный экстремум?
2. Сопоставьте задачи на условный и абсолютный экстремум.
3. Напишите функцию Лагранжа.
4. Сформулируйте необходимое условие локального условного экстремума (аналитическая форма).
5. Сформулируйте необходимое условие локального условного экстремума (геометрическая форма).
6. Приведите формулировку задачи математического программирования.
7. Приведите формулировку задачи линейного программирования.

ГЛАВА 9
МАКСИМИЗАЦИЯ ПОЛЕЗНОСТИ. ИССЛЕДОВАНИЕ МОДЕЛИ ПОТРЕБИТЕЛЬСКОГО СПРОСА. КОМПЕНСАЦИОННЫЕ ЭФФЕКТЫ

9.1. Функция полезности. Задача потребительского выбора

В данной главе будут рассмотрены некоторые модели потребительского выбора. Будем считать, что потребитель располагает доходом I, который он полностью тратит на приобретение благ (продуктов). Точнее говоря, величина I - это не доход, а расход данного потребителя. Потребитель решает статическую задачу, то есть в модели не учитываются его межвременные предпочтения и возможности делать или расходовать сбережения. Цены благ считаются заданными. Учитывая структуру цен, доход и собственные предпочтения, потребитель приобретает определенные количества благ, и математическая модель такого его поведения называется *моделью потребительского выбора*. Вначале мы рассмотрим модель с двумя видами благ. Такая модель удобна прежде всего возможностью графической интерпретации, сохраняя при этом все принципиальные свойства общей модели.

Рассмотрим потребительские наборы из двух благ. Потребительский набор (для краткости набор) - это вектор (x_1,x_2), координата x_1 которого равна количеству единиц первого блага, а координата x_2 равна количеству единиц второго блага.

Выбор потребителя (индивидуума) характеризуется отношением предпочтения, суть которого состоит в следующем. Считается, что потребитель про каждые 2 набора может сказать, что либо один из них более желателен, чем другой, либо потребитель не видит между ними разницы. Отношение предпочтения транзитивно, т.е. если набор $A=(a_1,a_2)$ предпочтительнее набора $B=(b_1,b_2)$, а набор B предпочтительнее набора $C=(c_1,c_2)$, то набор A предпочтительнее набора C.

На множестве потребительских наборов (x_1,x_2) определена функция $u(x_1,x_2)$ (называемая *функцией полезности потребителя)*, значение $u(x_1,x_2)$ которой на потребительском наборе (x_1,x_2) равно потребительской оценке индивидуума для этого набора. Потребительскую оценку $u(x_1,x_2)$ набора (x_1,x_2) принято называть *уровнем* (или *степенью*) удовлетворения потребностей индивидуума, если он приобретает или потребляет данный набор (x_1,x_2). Каждый потребитель имеет, вообще говоря, свою функцию полезности. Если набор A предпочтительнее набора B, то $u(A)>u(B)$.

Функция полезности удовлетворяет следующим свойствам:

1) Возрастание потребления одного продукта при постоянном потреблении другого продукта ведет к росту потребительской оценки, т.е.

$$\text{если } x_1^2 > x_1^1, \text{ то } u(x_1^2, x_2) > u(x_1^1, x_2);$$
$$\text{если } x_2^2 > x_2^1, \text{ то } u(x_1, x_2^2) > u(x_1, x_2^1).$$

1') Пусть $\dfrac{\partial u(x_1,x_2)}{\partial x_1} = u_1' > 0$, $\dfrac{\partial u(x_1,x_2)}{\partial x_2} = u_2' > 0$.

Из свойства 1') следует свойство 1).

Первые частные производные называются *предельными полезностями* продуктов: u_1' называется предельной полезностью первого продукта, u_2' - предельная полезность второго продукта. Для предельных полезностей первого и второго продуктов используется также символика $M_1u(x_1,x_2)$, $M_2u(x_1,x_2)$.

2) Предельная полезность каждого продукта уменьшается, если объем его потребления растет (это свойство предельной полезности называется *законом убывания предельной полезности).*

2') Пусть $\dfrac{\partial^2 u}{\partial x_1^2} = u_{11}'' < 0$, $\dfrac{\partial^2 u}{\partial x_2^2} = u_{22}'' < 0$.

Из свойства 2') следует свойство 2).

3) Предельная полезность каждого продукта увеличивается, если растет количество другого продукта. В этом случае продукт, количество которого фиксировано, оказывается относительно дефицитным. Поэтому дополнительная его единица приобретает большую ценность и может быть потреблена более эффективно. Данное свойство не столь очевидно, как 1)-2), и справедливо не для всех благ: если блага могут полностью замещать друг друга в потреблении, свойство 3) не выполняется. Предположение 3) вводится не всегда, но оно гарантирует выпуклость вниз линий безразличия.

3') Пусть $\dfrac{\partial^2 u}{\partial x_1 \partial x_2} = u_{12}'' = \dfrac{\partial^2 u}{\partial x_2 \partial x_1} = u_{21}'' > 0$.

Из свойства 3') следует свойство 3).

В учебной и монографической литературе понятие предельной полезности толкуется неоднозначно. Помимо приведенного выше определения предельной полезности первого (второго) продукта в виде частной производной u_1' (u_2') первого порядка, под предельной полезностью первого (второго) продукта понимают отношение приращения функции полезности к приращению вызвавшего его количества этого продукта:

$$M_1u(x_1,x_2) = \frac{u(x_1+\Delta x_1, x_2) - u(x_1,x_2)}{\Delta x_1} \quad \left(M_2u(x_1,x_2) = \frac{u(x_1, x_2+\Delta x_2) - u(x_1,x_2)}{\Delta x_2} \right).$$

Наконец, предельной полезностью первого (второго) продукта называют разность

$M_1u(x_1,x_2) = u(x_1+1,x_2) - u(x_1,x_2)$ ($M_2u(x_1,x_2) = u(x_1,x_2+1) - u(x_1,x_2)$)
или
$M_1u(x_1,x_2) = u(x_1,x_2) - u(x_1-1,x_2)$ ($M_2u(x_1,x_2) = u(x_1,x_2) - u(x_1,x_2-1)$).

Из контекста обычно бывает ясно, о каком конкретно толковании предельной полезности $M_1u(x_1,x_2)$ ($M_2u(x_1,x_2)$) идет речь.

Линия, соединяющая потребительские наборы (x_1,x_2), имеющие один и тот же уровень удовлетворения потребностей индивидуума, называется *линией безразличия*. Линия безразличия есть не что иное, как линия уровня функции полезности. Множество линий безразличия называется *картой* линий безразличия. На рис. 9.1 показан фрагмент карты линий безразличия. Линии безразличия, соответствующие разным уровням удовлетворения потребностей, не касаются и не пересекаются. Если линия безразличия $l_{\tau 3}$ расположена выше и правее ("северо-восточнее") линии безразличия то $\tau_3 > \tau_2$. Верно и обратное. Иными словами чем "северо-восточнее" расположена линия безразличия, тем большему уровню удовлетворения потребности она соответствует.

Условия 1)-3) означают, что линия безразличия убывает (является нисходящей) и строго выпукла к началу координат (к точке 0). Чтобы пояснить это, рассмотрим дифференциал (главную линейную часть приращения) функции $u(x_1,x_2)$. Если двигаться вдоль линии уровня, то приращение функции $u(x_1,x_2)$ равно нулю, и, следовательно, можно считать равной нулю и его главную линейную часть. Дифференциал функции полезности записывается следующим образом:

$$du(x_1,x_2)=u_1'dx_1+u_2'dx_2=0 \Rightarrow \frac{dx_2}{dx_1} = -\frac{u_1'}{u_2'} < 0. \qquad (1)$$

Итак, функция $x_2(x_1)$, то есть зависимость x_2 от x_1 вдоль кривой безразличия, является убывающей, поскольку производная ее отрицательна. Вторая производная функции $x_2(x_1)$ выглядит следующим образом:

$$d\left[\frac{dx_2}{dx_1}\right] \Big/ dx_1 = -\frac{u_{11}''\cdot u_2' - u_1'\cdot u_{21}''}{(u_2')^2} > 0. \qquad (2)$$

Ее положительность вытекает из свойств 1)-3); следовательно, кривые безразличия выпуклы вниз.

Рис. 9.1

Рассмотрим фиксированную линию безразличия l_τ. Пусть потребительский набор $(x_1,x_2) \in l_\tau$. При выполнении ряда естественных предположений (непрерывность первых частных производных u_1', u_2' и $u_2' \neq 0$) справедлива, как уже было показано, следующая формула:

$$\frac{dx_2}{dx_1} = -\frac{u_1'}{u_2'}. \tag{3}$$

Имеем приближенное равенство

$$\frac{dx_2}{dx_1} = -\mathrm{tg}\phi \approx -\mathrm{tg}\alpha = \frac{\Delta x_2}{\Delta x_1} \tag{4}$$

(см. рис. 9.2).

Из (3) и (4) следует важное приближенное равенство

$$-\frac{\Delta x_2}{\Delta x_1} \approx \frac{u_1'}{u_2'}. \tag{5}$$

Отношение $-\frac{\Delta x_2}{\Delta x_1}$ показывает, на сколько должен индивидуум увеличить (уменьшить) потребление *второго* продукта, если он уменьшил (увеличил) потребление *первого* продукта на одну единицу без изменения уровня удовлетворения своих потребностей (Это обстоя-

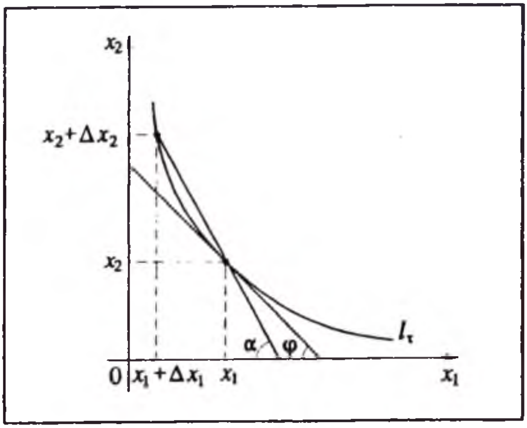


Рис. 9.2

тельство геометрически интерпретируется так: точки (x_1,x_2), $(x_1+\Delta x_1, x_2+\Delta x_2)$ принадлежат одной и той же линии безразличия l_r (см. рис. 9.2).) Поэтому дробь $-\frac{\Delta x_2}{\Delta x_1}$ принято называть *нормой замены* первого продукта вторым на потребительском наборе (x_1,x_2), а производную $-\frac{dx_2}{dx_1}$ (которая равна предельному значению дроби $-\frac{\Delta x_2}{\Delta x_1}$ при $\Delta x_1 \to 0$) - *предельной нормой замены* первого продукта вторым.

Примером функции полезности может служить функция

$$u(x_1,x_2)=a_1\log(x_1-\bar{x}_1)+a_2\log(x_2-\bar{x}_2), \tag{6}$$

где $a_1>0$, $a_2>0$, $x_1>\bar{x}_1\geq 0$, $x_2>\bar{x}_2\geq 0$.

Действительно, имеем $u_1'=\frac{a_1}{x_1-\bar{x}_1}>0$, $u_2'=\frac{a_2}{x_2-\bar{x}_2}>0$, $\frac{\partial^2 u}{\partial x_1^2}=-\frac{a_1}{(x_1-\bar{x}_1)^2}<0$, $\frac{\partial^2 u}{\partial x_2^2}=-\frac{a_2}{(x_2-\bar{x}_2)^2}<0$, т.е. выполнены свойства 1') и 2') функции полезности. Свойство 3') не выполнено, так как смешанные вторые частные производные функции $u(x_1,x_2)$ равны нулю.

Задача потребительского выбора (задача рационального поведения потребителя на рынке) заключается в выборе такого потребительского набора (x_1^0,x_2^0), который максимизирует его функцию полезности при заданном бюджетном ограничении.

Бюджетное ограничение означает, что денежные расходы на продукты не могут превышать денежного дохода, т.е. $p_1x_1 + p_2x_2 \le I$, где p_1 и p_2 - рыночные цены одной единицы первого и второго продуктов соответственно, а I - доход индивидуума, который он готов потратить на приобретение первого и второго продуктов. Величины p_1, p_2 и I заданы.

Формально задача потребительского выбора имеет вид:

$$\begin{gathered} u(x_1,x_2)\ (\max) \\ \text{при условиях (7)} \qquad \qquad (7) \\ p_1x_1 + p_2x_2 \le I, \\ x_1 \ge 0,\ x_2 \ge 0. \end{gathered}$$

Допустимое множество (то есть множество наборов благ, доступных для потребителя) представляет собой треугольник, ограниченный осями координат и бюджетной прямой (см. рис. 9.3). На этом множестве требуется найти точку, принадлежащую кривой безразличия с максимальным уровнем полезности. Поиск этой точки можно интерпретировать графически как последовательный переход на линии все более высокого уровня полезности (вправо-вверх) до тех пор, пока эти линии еще имеют общие точки с допустимым множеством.

9.2. Решение задачи потребительского выбора и его свойства

Набор (x_1^0, x_2^0), который является решением задачи потребительского выбора, принято называть *оптимальным* для потребителя, или *локальным рыночным равновесием* потребителя.

Вначале остановимся на некоторых важных свойствах задачи потребительского выбора. Во-первых, решение задачи (x_1^0, x_2^0) сохраняется при любом монотонном (то есть сохраняющем порядок значений) преобразовании функции полезности $u(x_1,x_2)$. Поскольку значение $u(x_1^0,x_2^0)$ было максимальным на всем допустимом множестве, оно остается таковым и после монотонного преобразования функции полезности (допустимое множество, определяемое бюджетным ограничением, остается неизменным). Таким монотонным преобразованием может быть умножение функции полезности на некоторое положительное число, возведение ее в положительную степень, логарифмирование по основанию, большему единицы. Отметим, что свойство 1) должно присутствовать у любой функции полезности; свойства 2) и 3) могут при ее монотонных преобразованиях теряться или приобретаться (рассмотрите это самостоятельно на примере функции $u(x_1,x_2)=x_1^{1/2}x_2^{1/2}$. Последнее важно для иллюстрации того факта, что если функция полезности в задаче потребительского выбора не обладает свойствами 2) или 3), это вовсе не означает, что данная задача не может описывать реальное поведение потребителя.

Во-вторых, решение задачи потребительского выбора не изменится, если все цены и доход увеличиваются (уменьшаются) в одно и то же число раз λ.

Это равнозначно умножению на положительное число λ обеих частей бюджетного ограничения $p_1x_1 + p_2x_2 \le I$, что дает неравенство, эквивалентное исходному. Поскольку ни цены, ни доход I не входят в функцию полезности, задача остается той же, что и первоначально.

В приведенной постановке задача потребительского выбора является задачей нелинейного программирования - см. главу 8, раздел 3. Однако если на каком-то потребительском наборе (x_1,x_2) бюджетное ограничение $p_1x_1+p_2x_2 \le I$ будет выполняться в виде строгого неравенства, то мы можем увеличить потребление какого-либо из продуктов и тем самым увеличить функцию полезности. Следовательно, набор (x_1^0,x_2^0), максимизирующий функцию полезности, должен обращать бюджетное ограничение в равенство, т.е. $p_1x_1^0+p_2x_2^0=I$. Графически это означает, что решение (x_1^0,x_2^0) задачи потребительского выбора должно лежать на бюджетной прямой (см. рис. 9.3), которую удобнее всего провести через точки пересечения с осями координат, где весь доход тратится на один продукт: $\left(0;\frac{I}{p_2}\right)$ и $\left(\frac{I}{p_1};0\right)$.

Мы также будем считать, что в оптимальной точке (x_1^0,x_2^0) условия $\{x_1 \ge 0,\ x_2 \ge 0\}$ выполняются автоматически, вытекая из свойств функции $u(x_1,x_2)$. Как правило, это действительно так. В то же время, если условия неотрицательности переменных не включать в явном виде в условие задачи, то она становится существенно проще с математической точки зрения.

Итак, задачу потребительского выбора можно заменить задачей на условный экстремум (ибо решение (x_1^0,x_2^0) этих двух задач одно и то же)

$$u(x_1,x_2)\ (\max)$$

при условии (8)

$$p_1x_1 + p_2x_2 = I$$

Для решения этой задачи на условный экстремум применим метод Лагранжа.

Выписываем функцию Лагранжа

$$L(x_1,x_2,\lambda) = u(x_1,x_2) + \lambda(p_1x_1 + p_2x_2 - I), \tag{9}$$

находим ее первые частные производные по переменным x_1, x_2 и и λ приравниваем эти частные производные к нулю: $\frac{\partial L}{\partial x_1}=u_1'-\lambda p_1=0$, $\frac{\partial L}{\partial y}=u_2'-\lambda p_2=0$, $\frac{\partial L}{\partial \lambda}=p_1x_1+p_2x_2-I=0$.

Исключив из полученной системы трех уравнений с тремя неизвестными неизвестную λ, получим систему двух уравнений с двумя неизвестными x_1 и x_2

$$\frac{u_1'}{u_2'} = \frac{p_1}{p_2},$$
$$p_1x_1+p_2x_2=I.$$

Решение (x_1^0,x_2^0) этой системы есть "укороченная" критическая точка функции Лагранжа. Можно строго доказать, что "укороченная" критическая точка (x_1^0,x_2^0) функции Лагранжа обязательно есть решение задачи потребительского выбора (за исключением так называемых угловых решений, которые здесь не рассматриваются). Подставив решение (x_1^0,x_2^0) в левую часть равенства

$$\frac{u_1'(x_1,x_2)}{u_2'(x_1,x_2)} = \frac{p_1}{p_2},$$

получим, что в точке (x_1^0,x_2^0) локального рыночного равновесия индивидуума отношение $\frac{u_1'(x_1^0,x_2^0)}{u_2'(x_1^0,x_2^0)}$ предельных полезностей $u_1'(x_1^0,x_2^0)$ и $u_2'(x_1^0,x_2^0)$ продуктов равно отношению рыночных цен p_1 и p_2 на эти продукты:

$$\frac{u_1'(x_1^0,x_2^0)}{u_2'(x_1^0,x_2^0)}=\frac{p_1}{p_2}. \qquad (10)$$

В связи с тем, что отношение $\frac{u_1'(x_1^0,x_2^0)}{u_2'(x_1^0,x_2^0)}$ равно предельной норме замены первого продукта вторым *в точке локального рыночного равновесия* (x_1^0,x_2^0), из (10) следует, что эта предельная норма равна отношению рыночных цен $\frac{p_1}{p_2}$ на продукты. Приведенный результат играет важную роль в экономической теории.

Геометрически решение (x_1^0,x_2^0) можно интерпретировать как точку касания линии безразличия функции полезности $u(x_1,x_2)$ с бюд-

Рис. 9.3

жетной прямой $p_1x_1+p_2x_2=I$ (см. рис. 9.3). Это определяется тем, что отношение $\frac{dx_2}{dx_1} = -\frac{u_1'}{u_2'}$ показывает тангенс угла наклона линии уровня функции полезности, а отношение $-\frac{p_1}{p_2}$ представляет тангенс угла наклона бюджетной прямой. Поскольку в точке потребительского выбора (или локального рыночного равновесия) они равны, в этой точке происходит касание данных двух линий.

Из (5) и (10) следует, что

$$-\frac{\Delta x_2^0}{\Delta x_1^0} \approx \frac{p_1}{p_2} \tag{11}$$

т. е. отношение (со знаком минус) конечных (относительно небольших) изменений Δx_2^0 и Δx_1^0 объемов продуктов в *локальном рыночном равновесии* (x_1^0,x_2^0) приближенно равно отношению рыночных цен p_1 и p_2 на продукты.

Равенство (11) позволяет давать приближенные оценки отношению рыночных цен, если известны конечные изменения объемов продуктов относительно потребительского набора, приобретенного потребителем, т. е. набора, который естественно следует толковать в качестве оптимального для потребителя.

Координаты x_1^0 и x_2^0 решения (x_1^0,x_2^0) задачи потребительского выбора есть функции параметров p_1,p_2 и I:

$$x_1^0 = x_1^0(p_1, p_2, I),$$
$$x_2^0 = x_2^0(p_1, p_2, I).$$

Полученные функции называются *функциями спроса* на первый и второй продукты. Важным свойством функций спроса является их однородность нулевой степени относительно цен и дохода, т.е. значения функций спроса инвариантны по отношению к пропорциональным изменениям цен и дохода.

$$x_1^0(\alpha p_1, \alpha p_2, \alpha I) = x_1^0(p_1, p_2, I)$$
$$x_2^0(\alpha p_1, \alpha p_2, \alpha I) = x_2^0(p_1, p_2, I)$$

для любого числа $\alpha > 0$. Это означает, что если все цены и доход изменятся в одно и то же число раз, величина спроса на продукт (первый или второй - безразлично) останется неизменной.

Пример задачи потребительского выбора

Решим одну простую задачу потребительского выбора с двумя благами. Пусть неизвестные количества этих благ равны x_1 и x_2, а их рыночные цены p_1 и p_2:

$$\begin{gathered} U(x_1, x_2) = x_1 x_2 \ (\max), \\ p_1 x_1 + p_2 x_2 \le I, \\ x_1 \ge 0,\ x_2 \ge 0. \end{gathered} \tag{12}$$

Как мы выяснили, бюджетное ограничение в оптимальной точке должно выполняться как равенство, и, поскольку оба блага жизненно необходимы (полезность равна нулю, если одно из них отсутствует), требования неотрицательности переменных будут выполнены автоматически. Следовательно, решаемая задача математического программирования превращается в классическую задачу на условный экстремум. Записав необходимые условия экстремума (согласно которым, отношения предельных полезностей благ должны равняться отношениям их рыночных цен, а бюджетное ограничение выполняется как равенство), получаем систему уравнений:

$$\frac{x_2}{x_1} = \frac{p_1}{p_2},\ p_1 x_1 + p_2 x_2 = I.$$

Первое условие означает, что в рассматриваемой задаче количества денег, затрачиваемые на оба блага, должны быть одинаковыми, то есть $x_2 p_2 = x_1 p_1$. Это вытекает из равенства “весов”, или показателей степени у переменных x_1 и x_2 в функции полезности. Итак, $x_2 p_2 + x_1 p_1 = \frac{I}{2}$ и функции спроса приобретают вид

$$x_1 = \frac{I}{2 \cdot p_1};\ x_2 = \frac{I}{2 \cdot p_2}. \qquad (13)$$

Итак, расход на каждое благо составляет половину общего дохода потребителя, и чтобы найти необходимое количество каждого блага, следует разделить расходуемую на него сумму на его цену.

Для этой простой модели мы могли бы найти решение без использования метода множителей Лагранжа, выражая x_2 через x_1 из бюджетного ограничения, подставляя это выражение в функцию полезности (которая становится полиномом второй степени от одной переменной) и находя максимум полученной квадратичной функции. Проделайте это как самостоятельное упражнение, получив те же самые функции спроса. Для более сложных случаев, некоторые из которых будут рассмотрены в следующей главе, решить задачу элементарными методами сложно, и требуются методы дифференциального исчисления (например, тот же метод множителей Лагранжа) или математического программирования.

Случай потребительских наборов $(x_1, ..., x_n)$ из n продуктов принципиально не отличается от случая двух продуктов, но технически несколько сложнее. Этот случай здесь отдельно не рассматривается.

9.3. Общая модель потребительского выбора

В предыдущем разделе рассмотрена типовая модель потребительского выбора с двумя товарами и ее решение с помощью метода множителей Лагранжа. Сейчас мы рассмотрим свойства задачи потребительского выбора с произвольным числом товаров и целевой функцией общего вида, а затем перейдем к некоторым конкретным задачам, включая анализ компенсированного изменения цен.

Пусть задана целевая функция предпочтения потребителя $u(x_1, x_2, \ldots, x_n)$, где x_i - количество i-го блага, вектор цен $\{p_i\} = (p_1, p_2, ..., p_n)$ и доход I. Записав бюджетное ограничение и ограничения на неотрицательность, получаем задачу

$$u(x) \to \max \qquad (14)$$

при условиях

$$px \le I,\ x \ge 0 \text{ (здесь } x=(x_1,\ldots,x_n),\ p=(p_1,\ldots,p_n),\ px=p_1x_1+\ldots+p_nx_n).$$

Будем, как и ранее, считать, что неотрицательность переменных обеспечивается свойствами целевой функции и бюджетного ограничения. В этом случае можно записать функцию Лагранжа и исследовать ее на безусловный экстремум.

Функция Лагранжа $L(x,\lambda) = u(x) + \lambda(px - I)$.

Необходимые условия экстремума - равенство нулю частных производных:

$L_i' = u_i' + \lambda p_i = 0$ для всех i от единицы до n и $L_\lambda' = px - I = 0$.

Отсюда вытекает, что для всех i, j в точке x^0 локального рыночного равновесия выполняется равенство

$$\frac{u_i'}{u_j'} = \frac{p_i}{p_j} \tag{15}$$

которое получается после перенесения вторых слагаемых необходимых условий в правую часть и делением i-го равенства на j-ое. Итак, в точке оптимума отношение предельных полезностей любых двух благ равно отношению их рыночных цен. Равенство (15) можно переписать и в другой форме:

$$\frac{u_i'}{p_i} = \frac{u_j'}{p_j} \tag{16}$$

Последнее означает, что дополнительная полезность, приходящаяся на дополнительную единицу денежных затрат, в точке оптимума одинакова по всем видам благ. Если бы это было не так, то по крайней мере одну денежную единицу можно было бы перераспределить так, чтобы выросло благосостояние (значение функции полезности) потребителя. Если для некоторых i,j $\frac{u_i'}{p_i} > \frac{u_j'}{p_j}$, то некоторое количество денег можно было бы перераспределить от i-го блага к j-му, увеличив уровень благосостояния.

9.3.1. Модель Р.Стоуна

Выведем теперь функцию спроса для конкретной функции потребительского предпочтения, называемой функцией Р.Стоуна. Эта функция имеет вид:

$$u(x) = \prod_{i=1}^{n} (x_i - a_i)^{\alpha_i} \to \max. \tag{17}$$

Здесь a_i - минимально необходимое количество i-го блага, которое приобретается в любом случае и не является предметом выбора. Для того чтобы набор $\{a_i\}$ мог быть полностью приобретен, необходимо, чтобы доход I был больше $\sum_i p_i a_i$ - количества денег, необходимого для покупки этого набора. Коэффициенты степени $\alpha_i>0$ характеризуют относительную "ценность" благ для потребителя.

Добавив к целевой функции (17) бюджетные ограничения $\sum_i p_i a_i \le I$, $x_1 \ge 0,\dots, x_n \ge 0$, получим задачу, называемую моделью Р.Стоуна. Приравняв нулю частные производные функции Лагранжа по переменным x_i, получаем для всех i от 1 до n: $\frac{\alpha_i u(x)}{x_i - a_i} + \lambda p_i = 0$, откуда

$$x_i = a_i - \frac{\alpha_i u(x)}{\lambda p_i}. \tag{18}$$

К этим условиям добавляется равенство $\sum_i p_i x_i - I = 0$, выполнение которого эквивалентно равенству нулю частной производной функции Лагранжа по переменной λ. Умножив каждое i-е условие на λp_i и просуммировав их по i, имеем

$$\sum_i \alpha_i u(x)_i + \lambda \sum_i p_i x_i - \lambda \sum_i p_i a_i = 0. \tag{19}$$

Поскольку в точке оптимума бюджетное ограничение выполняется как равенство, заменим $\sum_i p_i x_i$ на I. Получим $\frac{u(x)}{\lambda} = -\frac{I - \sum p_i a_i}{\sum \alpha_i}$.

Отсюда имеем функцию спроса:

$$x_i = a_i + \frac{\alpha_i \left(I - \sum_j p_j a_j\right)}{p_i \sum_j \alpha_j} \tag{20}$$

Эту функцию легко проинтерпретировать и запомнить следующим образом. Вначале приобретается минимально необходимое количество каждого блага a_i. Затем рассчитывается сумма денег, остающаяся после этого, которая распределяется пропорционально "весам" важности α_i. Разделив количество денег на цену p_i, получаем дополнительно приобретаемое, сверх минимума, количество i-го блага и добавляем его к a_i.

Модель Стоуна имеет различные частные случаи: например, когда все $a_i = 0$, а все α_i равны между собой, получаем $x_i = \frac{I}{np_i}$ (то есть доход делится на n равных частей и спрос на i-й товар рассчитывается как частное от деления полученной суммы денег на его цену).

В данном случае мы видим, что спрос растет при росте дохода с эластичностью, равной единице, и уменьшается с ростом цены с эластичностью, равной минус единице. Тем самым каждый товар в этой модели является нормальным и ценным. Кроме того, спрос растет до бесконечности при бесконечном росте дохода - в этом смысле каждый товар является “предметом роскоши”.

Для того чтобы описать более разнообразные формы поведения спроса на различные товары, модель должна включать другие, более сложные виды целевой функции предпочтения. Например, при функции предпочтения

$$u(x_1,x_2)=x_1^a x_2^{b-a}(x_1 + b - a)^{-b}$$

(где a, b - параметры) функция спроса имеет вид

$x_1 = \dfrac{aI}{I + bp_1}$ (типичная функция спроса для предметов первой необходимости) и $x_2 = \dfrac{I(I + p_1(b - a))}{I + bp_1}$ (типичная функция спроса для предметов роскоши).

9.3.2. Взаимозаменяемость благ. Эффекты компенсации

Если функция спроса имеет вид $x_i = \dfrac{I}{np_i}$ (или, при не равных между собой α_i, $x_i = \dfrac{I\alpha_i}{p_i\sum_j \alpha_j}$), то спрос на i-й товар не зависит от цены на любой j-й товар. Вообще говоря, перекрестные функции спроса от цен характеризуют такие свойства товаров, как взаимозаменяемость и взаимодополняемость. Если при росте цены на товар i, при снижении спроса на i-й товар, растет спрос на товар j - эти товары взаимозаменяемы. Наоборот, если спрос на j-й товар также падает, - они взаимодополняемы. Заметим, что реальная взаимозаменяемость может искажаться общим снижением благосостояния при росте цены i-го блага: j-е благо может заменять i-е в потреблении, но спрос на него может не расти, поскольку снизилось общее благосостояние потребителя. Для снятия этого искажения используют понятие компенсированного изменения цены, то есть такого, которое сопровождается увеличением дохода потребителя, позволяющим ему поддерживать прежний уровень благосостояния. Практически компенсированное изменение цены изображается следующим образом (рис. 9.4).

Рис. 9.4

Пусть цена первого блага повысилась с p_1^1 до p_1^2, тогда бюджетная прямая из положения 1 перейдет в положение 2. Точка A на линии безразличия $I_{\tau 1}$, касающейся первоначального бюджетного ограничения, будет заменена новой точкой оптимума B, где новая линия безразличия $I_{\tau 2}$ касается новой бюджетной прямой. Если мы хотим *компенсировать* потребителю потерю благосостояния, то увеличим его доход так, чтобы новая бюджетная прямая 3 (параллельная линии 2) коснулась в некоторой точке C прежней линии безразличия $I_{\tau 1}$. Направленный отрезок AC показывает "эффект замены" при росте цены, то есть изменение структуры спроса при условии поддержания прежнего уровня благосостояния. Направленный отрезок CB отражает "эффект дохода", то есть изменение потребительского спроса при сохранении соотношения цен благ и изменении уровня дохода. Общий результат роста цены (при отсутствии компенсации) выражается направленным отрезком AB.

Для формального анализа компенсационных эффектов рассмотрим вначале две задачи.

Пусть целевая функция потребителя (ЦФП) зависит от двух благ, x_1 и x_2, следующим образом: $u(x_1,x_2) = x_1 x_2 \to \max$. Пусть цены благ равны, соответственно, 10 и 2, а доход потребителя - 60. Тогда, согласно полученной формуле функции спроса, $x_1=\frac{60}{2\cdot 10}=3$; $x_2=\frac{60}{2\cdot 2}=15$; $u^*=45$. Пусть теперь p_2 меняется с 2 до 7. Каков необходимый размер компенсации? Чтобы приобрести прежний оптимальный набор, потребителю необходимо дополнительно (7-2)·15 = 75 денежных единиц. Однако прежняя структура потребления не будет оптимальной при новых ценах, и минимальная необходимая компенсация будет меньше, чем 75. Пусть потребитель получает дополнительно количество денег M. Тогда при новых ценах его спрос на первое и второе блага будет равен: $x_1 = \frac{60 + M}{10 \cdot 2}$; $x_2 = \frac{60 + M}{7 \cdot 2}$. Целевая функция $x_1 \cdot x_2$ будет равна $\frac{(60 + M)^2}{10 \cdot 7 \cdot 4}$, и это выражение должно равняться начальному u^*=45. Отсюда $M \approx 52{,}25$, что существенно меньше, чем 75.

Теперь решим задачу в более общем виде. Пусть по-прежнему $u(x_1,x_2) = x_1 x_2$, цены благ равны p_1 и p_2, а доход I. Очевидно,

$$x_i = \frac{I}{2p_i};\ \frac{\partial x_i}{\partial p_i} = -\frac{I}{2p_i^2};\ \frac{\partial x_i}{\partial I} = \frac{1}{2p_i};\ \frac{\partial x_i}{\partial p_j} = 0$$

Пусть теперь цена p_1 выросла в z раз ($z > 1$), и при этом потребитель получает необходимую компенсацию. Новый размер дохода обозначим через $\bar{I}$, спрос - $\bar{x}_1$ и $\bar{x}_2$. Очевидно, $\bar{x}_1 = \frac{\bar{I}}{2zp_1}$; $\bar{x}_2 = \frac{\bar{I}}{2p_2}$ и условие компенсации $\frac{\bar{I}^2}{4zp_1p_2} = \frac{I^2}{4p_1p_2}$, откуда $\bar{I} = \sqrt{z} \cdot I$; $\bar{x}_1 = \frac{x_1}{\sqrt{z}}$; $\bar{x}_2 = x_2\sqrt{z}$.

Итак, спрос на первый товар в случае с компенсацией сократится в $\sqrt{z}$ раз (а не в z раз, как без нее), а спрос на второй товар в $\sqrt{z}$ раз вырастет. В случае роста цены второго товара ситуация будет полностью симметричной. Таким образом, $\left.\left(\frac{\partial x_i}{\partial p_j}\right)\right|_{comp} > 0$ при $i = 1$, $j = 2$ или $j = 1$, $i = 2$. Индекс *comp* означает, что перекрестная частная производная спроса рассчитывается при необходимой для поддержания прежнего уровня благосостояния компенсации дохода. Условие компенсации снимает "эффект дохода", оставляя лишь"эффект замены", что позволяет более точно определить понятие взаимозаменяемости и взаимодополняемости благ и оценивать эти характеристики. Блага i и j называются взаимозаменяемыми, если $\left.\left(\frac{\partial x_i}{\partial p_j}\right)\right|_{comp} > 0$ и $\left.\left(\frac{\partial x_j}{\partial p_i}\right)\right|_{comp} > 0$ (эти два условия равносильны), и взаимодополняемыми, если $\left.\left(\frac{\partial x_i}{\partial p_j}\right)\right|_{comp} < 0$ и $\left.\left(\frac{\partial x_j}{\partial p_i}\right)\right|_{comp} < 0$.

Рассчитаем теперь эти частные производные для рассматриваемой задачи, когда p_1 растет в z раз. В этом случае приращение $\Delta x_1 = \frac{x_1}{\sqrt{z}} - x_1$; $\Delta x_2 = \sqrt{z}\, x_2 - x_2$; $\Delta p_1 = z p_1 - p_1$. Отсюда

$$\left[\frac{\partial x_1}{\partial p_1}\right]_{comp} = \lim_{z \to 1} \frac{x_1(1 - \sqrt{z})}{p_1\sqrt{z}\cdot(z-1)} = \lim_{z \to 1}\left[- \frac{x_1}{p_1\sqrt{z}(1 + \sqrt{z})}\right] = - \frac{x_1}{2p_1} = - \frac{I}{4p_1^2}$$

$$\left[\frac{\partial x_2}{\partial p_1}\right]_{comp} = \lim_{z \to 1} \frac{x_2(\sqrt{z} - 1)}{p_1\sqrt{z}(z-1)} = \lim_{z \to 1} \frac{x_2}{p_1\sqrt{z}(\sqrt{z} + 1)} = \frac{x_2}{2p_2} = \frac{I}{4p_1p_2}.$$

Последняя величина положительна, что свидетельствует о взаимозаменяемости благ в рассматриваемой задаче.

9.3.3. Уравнение Слуцкого

Одним из основных в теории потребительского выбора является уравнение Слуцкого, опубликованное российским математиком Е.Е.Слуцким в 1915 году. Это уравнение позволяет увязать действие эффекта замены и эффекта дохода с результирующим изменением спроса. Мы не будем выводить уравнение Слуцкого, лишь приведем его в используемых здесь обозначениях, сделав некоторые комментарии:

$$\frac{\partial x_i}{\partial p_j} = \left[\frac{\partial x_i}{\partial p_j}\right]_{comp} - \left[\frac{\partial x_i}{\partial I}\right] x_j \qquad (21)$$

Первое слагаемое в правой части описывает действие эффекта замены, второе - действие эффекта дохода, выраженное в тех же единицах измерения (множитель x_j приводит их к одной размерности). Слева записано результирующее воздействие на спрос, складывающееся из изменения структуры спроса и общего его изменения при изменении уровня реального дохода. Для ценных товаров величина $\frac{\partial x_i}{\partial I} > 0$, т.е. спрос растет при росте дохода. В этом случае, согласно уравнению Слуцкого, $\frac{\partial x_i}{\partial p_j} < \left[\frac{\partial x_i}{\partial p_j}\right]_{comp}$: если спрос растет, то он растет больше при наличии компенсации, если падает - то в меньшей степени. Может оказаться и так, что $\frac{\partial x_i}{\partial p_j} < 0$, но $\left|\frac{\partial x_i}{\partial p_j}\right|_{comp} > 0$, то есть товары i и j взаимозаменяемы, но представляются взаимодополняемыми без учета компенсации. Уравнение Слуцкого может рассматриваться как при разных, так и при совпадающих i и j (запишите его для последнего случая самостоятельно). Из обычно постулируемых свойств функции полезности потребителя 1')-2') вытекает, что $\left|\frac{\partial x_i}{\partial p_i}\right|_{comp} < 0$ (на графике это обусловлено выпуклостью линий уровня функции полезности). Если, в таком случае, вдруг оказывается, что $\frac{\partial x_i}{\partial p_i} > 0$ (спрос на товар растет при росте цены - такие товары называются товарами Гиффина), то отсюда вытекает, что $\frac{\partial x_i}{\partial I} < 0$ - то есть это обязательно малоценный (худший) товар. Обычно приводимый в качестве примера товара Гиффина картофель удовлетворяет этому условию, в то же время золото, например, не может ему удовлетворять. Таким образом, рост спроса на золото при росте его цены, наблюдавшийся одно время в СССР, нужно объяснить другими причинами, - главным образом, скрытой инфляцией и узостью сферы вложения свободных средств населения в тот период.

Выпишем и проверим уравнение Слуцкого для рассмотренной выше задачи потребительского выбора с функцией полезности $u(x_1,x_2)=x_1 x_2$. Как было получено,

$$\frac{\partial x_1}{\partial p_1}=-\frac{1}{2p_1^2};\quad \frac{\partial x_i}{\partial I}=\frac{1}{2p_i};\quad \frac{\partial x_2}{\partial p_1}=0;\quad \left|\frac{\partial x_1}{\partial p_1}\right|_{comp}=-\frac{I}{4}p_1^2;\quad \left|\frac{\partial x_2}{\partial p_1}\right|_{comp}=\frac{I}{4}p_1p_2.$$

Отсюда

$$-\frac{I}{2p_1^2}=-\frac{I}{p_1^2}-\left|\frac{1}{2p_1}\right|\cdot\left|\frac{I}{2p_1}\right|=-\frac{I}{2p_1^2},\quad \text{и}\quad 0=\frac{I}{4p_1p_2}-\left|\frac{1}{2p_2}\right|\cdot\left|\frac{I}{2p_1}\right|=0.$$

Итак, в обоих случаях уравнения Слуцкого (при $i = j$ и при $i j$) здесь выполнены. Уравнение Слуцкого может быть использовано для нахождения $\left|\frac{\partial x_i}{\partial p_j}\right|_{comp}$, то есть для расчета эффекта замены и оценки взаимозаменяемости или взаимодополняемости благ, поскольку частные производные без компенсации рассчитываются значительно легче (как это было показано выше).

Рассмотрим теперь более подробно эластичности функции спроса. Эластичность спроса по цене равна $e_{ij} = \frac{\partial x_i}{\partial p_j} : \frac{x_i}{p_j}$, эластичность спроса по доходу $e_{iI} = \frac{\partial x_i}{\partial I} : \frac{x_i}{I}$. Для функции $x_i = \alpha_i \frac{I}{\sum_j \alpha_j p_i}$ эластичность $e_{ii} = -1$; $e_{ij} = 0$ $(i\, j)$; $e_{iI} = 1$ (как это было показано выше).

Как уже говорилось, если в функции спроса $x_i = x_i(p_1, p_2, \ldots, p_n, I)$ все цены и доход увеличить в одно и то же количество раз λ, то спрос x_i не изменится. Таким образом, $x_i(\lambda p, \lambda I) = \lambda^0 x_i(p, I) = x_i(p, I)$, то есть, функция спроса является однородной нулевой степени. Отсюда, согласно уравнению Эйлера должно выполняться равенство

$$\sum_j \left[\frac{\partial x_i}{\partial p_j}\right] p_j + \left[\frac{\partial x_i}{\partial I}\right] I = 0 \tag{22}$$

разделив которое на x_i, получим равенство $\sum_j e_{ij} + e_{iI} = 0$, то есть нулю должна равняться сумма всех эластичностей спроса по ценам и доходу.

В качестве иллюстрации покажем, что если в задаче потребительского выбора всего два товара, то они обязательно являются взаимозаменяемыми. Для этого воспользуемся тем, что $\left[\frac{\partial x_i}{\partial p_i}\right]_{comp} < 0$, и положительностью частных производных функции полезности.

Предположим, что выросла цена 1-го товара p_1. Поскольку $\left[\frac{\partial x_1}{\partial p_1}\right]_{comp} < 0$, спрос на этот товар при условии компенсации падает. Если бы при этом упал спрос и на второй товар, то мы получили бы точку, в которой обоих товаров меньше, чем в начальной. Следовательно, в этой точке значение функции полезности $u(x_1, x_2)$ должно быть также меньше (а мы знаем, что в условиях компенсации оно равно начальному). Следовательно, спрос на второй товар при условии компенсации должен вырасти (т.е. $\left[\frac{\partial x_2}{\partial p_1}\right]_{comp} > 0$), и он является взаимозаменяемым с первым товаром.

Вопросы и задачи к главе 9

1. Что означает отношение предпочтения?
2. Каким свойствам должна удовлетворять функция полезности?
3. Каков экономический смысл свойств функции полезности?
4. Приведите пример функции полезности.
5. Сформулируйте задачу потребительского выбора.
6. Выпишите необходимые условия решения задачи потребительского выбора.
7. Приведите геометрическую интерпретацию решения задачи потребительского выбора.
8. Что такое функции спроса? В чем состоит условие их однородности нулевой степени, его экономический смысл?
9. Почему в точке оптимума задачи потребительского выбора бюджетное ограничение выполняется как равенство?
10. Изобразите графически линии уровня ЦФП и бюджетное ограничение так, чтобы ограничения $x_i \geq 0$ стали существенными для решения задачи потребительского выбора.
11. В точке оптимума полезности приращения благ, приходящиеся на одну затрачиваемую денежную единицу, равны между собой. Поясните.
12. Какие параметры потребительских предпочтений задаются экзогенно в модели Стоуна?
13. Запишите формулу для суммы денег, затрачиваемой для приобретения i-го товара в решении модели Стоуна.
14. Зависит ли сумма денег, расходуемая на товар i в решении модели Стоуна, от цены товара j ($i \neq j$):
    а) при $a_j = 0$?
    б) при $a_j > 0$?
15. Ведут ли себя товары в модели Стоуна как предметы первой необходимости или предметы роскоши с точки зрения зависимости спроса на них от дохода?
16. В чем состоит воздействие на спрос эффекта замены и эффекта дохода при изменении цены одного из благ? Изобразите графически семейства линий уровня ЦФП и бюджетное ограничение, когда эффекты замены и дохода воздействуют на спрос на некоторый товар:
    а) в одном направлении;
    б) в разных направлениях.
17. Равнозначно ли воздействие на потребительский спрос увеличение дохода в k раз и сокращение в k раз всех цен? Сделайте выводы для рассматриваемой модели и для реальности и сопоставьте их.

18. Предположим, что функция спроса на товар x_1 зависит от его цены p_1 и дохода потребителя I следующим образом: $\frac{2}{3}\frac{\sqrt{I}}{p_1^2}$. Используя уравнение Слуцкого, рассчитайте $\left|\frac{\partial x_1}{\partial p_1}\right|_{comp}$.
19. Могут ли все эластичности спроса на товар i по доходу и по ценам быть неотрицательными?
20. Запишите какую-нибудь функцию спроса, для которой спрос на i-й товар эластичен как по цене, так и по доходу, то есть $e_{ii} < -1$, а $e_{iI} > 1$.
21. Решите задачу потребительского выбора, найдя функции спроса, при ценах благ $p_1 = 10$, $p_2 = 2$ и доходе $I = 60$, со следующими функциями предпочтения:

   1) $u = x_1 x_2 \to \max$;
   2) $u = x_1^{1/2} x_2^{2/5} \to \max$;
   3) $u = (x_1-1)^{1/4} (x_2-3)^{3/4} \to \max$;
   4) $u = 5(4 - x_1)^2 + (20 - x_2)^2 \to \min$.

   Для каждой задачи изобразите допустимое множество и кривые безразличия.

ГЛАВА 10
ПРОИЗВОДСТВЕННЫЕ ФУНКЦИИ

10.1. Понятие производственной функции одной переменной

Производственная функция - это функция, независимая переменная которой принимает значения объемов *затрачиваемого* или *используемого ресурса* (фактора производства), а зависимая переменная - значения объемов *выпускаемой продукции*

$$y = f(x) \tag{1}$$

В формуле (1) x ($x \geq 0$) и y ($y \geq 0$) - числовые величины, т.е. $y = f(x)$ есть функция одной переменной x. В связи с этим производственная функция (ПФ) f называется одноресурсной или *однофакторной* ПФ, ее область определения - множество неотрицательных действительных чисел (т.е. $x \geq 0$). Запись $y = f(x)$ означает, что если ресурс затрачивается или используется в количестве x единиц, то продукция выпускается в количестве $y = f(x)$ единиц. Символ f - знак функции - является характеристикой *производственной системы*, преобразующей ресурс в выпуск. Символ f связывает между собой независимую переменную x с зависимой переменной y. В микроэкономической теории принято считать, что y - это *максимально* возможный объем выпуска продукции, если ресурс затрачивается или используется в количестве x единиц. В макроэкономике такое понимание не совсем корректно: возможно, при другом распределении ресурсов между структурными единицами экономики выпуск мог бы быть и большим. В этом случае ПФ - это статистически устойчивая связь между затратами ресурса и выпуском. Более правильной является символика $y = f(x, a)$, где a - вектор *параметров* ПФ.

Пример 1.1. Возьмем ПФ f в виде $f(x) = ax^b$, где x - величина затрачиваемого ресурса (например, рабочего времени), $f(x)$ - объем выпускаемой продукции (например, число готовых к отправке холодильников). Величины a и b - *параметры* ПФ f. Здесь a и b - положительные числа и число $b \leq 1$, вектор параметров есть двумерный вектор (a, b).

График Γ производственной функции $y = ax^b$ изображен на рис. 10.1. На графике видно, что с ростом величины затрачиваемого ресурса x объем выпуска y растет, однако при этом каждая *дополни-*

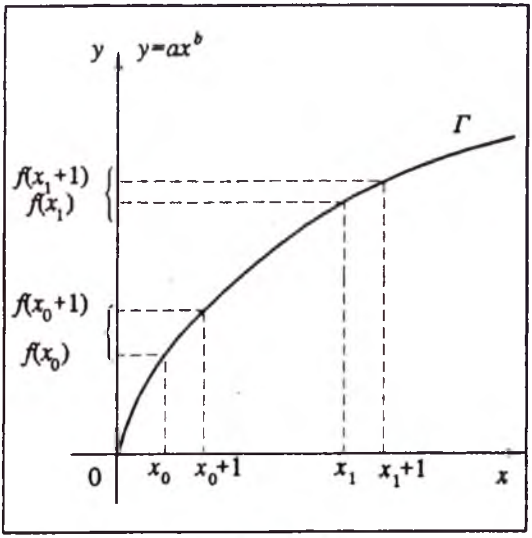


Рис. 10.1

тельная единица ресурса дает ***все меньший прирост*** объема y выпускаемой продукции. Отмеченное обстоятельство *(рост* объема y и *уменьшение* прироста объема y с ростом величины x) отражает фундаментальное положение экономической *теории* (хорошо подтверждаемое *практикой*), называемое ***законом убывающей эффективности***. ПФ $y = ax^b$ является типичным представителем широкого класса однофакторных ПФ.

ПФ могут иметь разные *области использования*. Принцип "затраты - выпуск" может быть реализован как на *микро-*, так и на *макроэкономическом* уровне. Сначала остановимся на микроэкономическом уровне. ПФ $y = ax^b$, рассмотренная выше, может быть использована для описания взаимосвязи между величиной затрачиваемого или используемого ресурса x в течение года на отдельном предприятии (фирме) и годовым выпуском продукции y этого предприятия (фирмы). В роли производственной системы здесь выступает *отдельное предприятие (фирма)* - имеем *микроэкономическую* ПФ (МИПФ). На микроэкономическом уровне в роли *производственной системы* может выступать также *отрасль, межотраслевой производственный комплекс*. МИПФ строятся и используются в основном для решения *задач анализа и планирования*, а также *задач прогнозирования*.

ПФ может быть использована для описания взаимосвязи между годовыми затратами труда в масштабе региона или страны в целом и годовым конечным выпуском продукции (или доходом) этого региона или страны в целом. Здесь в роли производственной системы выступает *регион* или *страна* в целом (точнее *хозяйственная система* региона или страны) - имеем *макроэкономический* уровень и *макроэкономическую* ПФ (МАПФ). МАПФ строятся и активно используются для решения всех трех типов задач (анализа, планирования и прогнозирования).

Точное толкование понятий затрачиваемого (или используемого) ресурса и выпускаемой продукции, а также выбор *единиц* их *измерения* зависят от характера и масштаба *производственной системы*, особенностей решаемых (с помощью ПФ) *задач* (аналитических, плановых, прогнозных), наличия *исходных данных*. На *микроэкономическом* уровне затраты и выпуск могут измеряться как в *натуральных*, так и в *стоимостных* единицах *(показателях)*. Годовые затраты труда могут быть измерены в человеко-часах (объем человеко-часов - натуральный показатель) или в рублях выплаченной заработной платы (ее величина - стоимостный показатель); выпуск продукции может быть представлен в штуках или в других натуральных единицах (тоннах, метрах и т.п.) или в виде своей стоимости.

На *макроэкономическом* уровне затраты и выпуск измеряются, как правило, в стоимостных показателях и представляют собой *стоимостные (ценностные) агрегаты* , т.е. суммарные величины произведений объемов затрачиваемых (или используемых) ресурсов и выпускаемых продуктов на их цены.

Производственная функция нескольких переменных - это функция, независимые переменные $x_1, \dots, x_n$ которой принимают значения объемов затрачиваемых или используемых ресурсов (число переменных n равно числу ресурсов), а значение функции имеет смысл величин объемов выпуска:

$$y = f(x) = f(x_1, \dots, x_n). \tag{2}$$

В формуле (2) y ($y \geq 0$) - скалярная, а x - векторная величина, $x_1, \dots, x_n$ - координаты вектора x, т.е. $f(x_1, \dots, x_n)$ есть числовая функция нескольких *(многих)* переменных $x_1, \dots, x_n$. В связи с этим ПФ $f(x_1, \dots, x_n)$ называют многоресурсной или *многофакторной* ПФ. Более правильной является такая символика $f(x_1, \dots, x_n, a)$, где a - вектор *параметров* ПФ.

По экономическому смыслу $x_1 \geq 0, \dots, x_n \geq 0$, следовательно, областью определения многофакторной ПФ $f(x_1, \dots, x_n)$ является множество n-мерных векторов x, все координаты $x_1, \dots, x_n$ которых неотрицательные числа.

Для отдельного предприятия (фирмы), выпускающего однородный продукт, ПФ $f(x_1, \dots, x_n)$ может связывать объем выпуска (в натуральном или стоимостном выражении) с затратами рабочего времени по различным видам трудовой деятельности, различных видов сырья, комплектующих изделий, энергии, основного капитала (измеренных обычно в натуральных единицах). ПФ такого типа характеризуют действующую *технологию* предприятия (фирмы).

При построении ПФ для региона или страны в целом в качестве величины годового выпуска Y (будем обозначать объем выпуска, или дохода, на макроуровне большой буквой) чаще берут совокупный продукт (доход) региона или страны, исчисляемый обычно в *неизменных,* а не в текущих *ценах,* в качестве ресурсов рассматривают *основной капитал* (x_1 ($=K$) - объем *используемого* в течение года основного капитала), *живой труд* (x_2 ($=L$) - количество единиц *затрачиваемого* в течение года живого труда), исчисляемые обычно в стоимостном выражении. Таким образом строят двухфакторную ПФ $f(x_1, x_2)$, или $Y = f(K,L)$. От двухфакторных ПФ переходят к трехфакторным. В качестве третьего фактора иногда вводят объемы используемых природных ресурсов. Кроме того, если ПФ строится по данным временных рядов, то в качестве особого фактора роста производства может быть включен технический прогресс.

ПФ $y = f(x_1, x_2)$ называется *статической,* если ее параметры и ее характеристика *f не зависят от времени t*, хотя объемы ресурсов и объем выпуска могут зависеть от времени t, т.е. могут иметь представление в виде *временных рядов:* $x_1(0), x_1(1), \ldots, x_1(T); x_2(0), x_2(1), \ldots, x_2(T); y(0), y(1), \ldots, y(T); y(t) = f(x_1(t), x_2(t))$. Здесь t - номер года, $t = 0, 1, \ldots, T$; $t = 0$ - базовый год временного промежутка, охватывающего годы 1, 2, ... T.

Пример 1.2. Для моделирования отдельного региона или страны в целом (т.е. для решения задач на макроэкономическом, а также и на микроэкономическом уровне) часто используется ПФ вида $y = a_0 x_1^{a_1} x_2^{a_2}$, где a_0, a_1, a_2 - параметры ПФ. Это положительные постоянные (часто a_1 и a_2 таковы, что $a_1 + a_2 = 1$). ПФ только что приведенного вида называется ПФ Кобба-Дугласа (ПФКД) по имени двух американских экономистов, предложивших ее использовать в 1929 г. ПФКД активно применяется для решения разнообразных теоретических и прикладных задач благодаря своей *структурной простоте.* ПФКД принадлежит к классу так называемых *мультипликативных* ПФ (МПФ). В приложениях ПФКД $x_1 = K$ равно объему *используемого* основного капитала (объему используемых основных фондов - в отечественной терминологии), $x_2 = L$ - *затратам* живого труда, тогда ПФКД приобретает вид, часто используемый в литературе:

$$Y = a_0 K^{a_1} L^{a_2}.$$

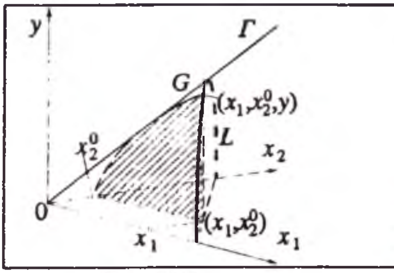


Рис. 10.2

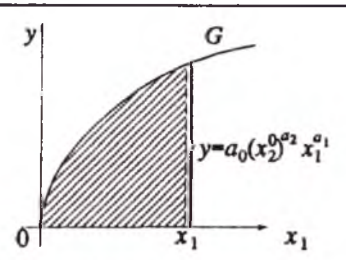


Рис. 10.3

Графиком ПФ $y = a_0 x_1^{a_1} x_2^{a_2}$ $(a_1 + a_2 = 1)$ в трехмерном пространстве является двумерная поверхность Γ, эскиз которой представлен на рис. 10.2. График Γ в рассматриваемом случае есть коническая поверхность, направляющей которой является, например, линия L, а образующими - лучи, выходящие из точки O. Пусть $x_2 = x_2^0 > 0$, тогда $y = (a_0(x_2^0)^{a_2}) x_1^{a_1}$ и мы получаем вариант ПФ, аналогичный рассмотренному выше (см. рис. 10.1 и рис. 10.3). Линия G есть пересечение поверхности Γ вертикальной плоскостью $x_2 = x_2^0$. На рис. 10.3 представлен фрагмент рис. 10.2, относящийся к линии G. Поведение линии G отражает то обстоятельство, что с ростом затрат первого ресурса объем выпуска y растет, но каждая дополнительная единица первого ресурса обеспечивает все меньший прирост выпуска y. Это обстоятельство можно прокомментировать следующим образом. Если число работников и их квалификация остаются неизменными, а число обслуживаемых ими станков (которое уже достаточно велико) увеличивается, например, в два раза, то это естественно *не* приведет к двойному росту объема выпуска. Отметим, что если $a_1 + a_2 < 1$, то графиком Γ ПФКД является поверхность, которая напоминает выпуклую вверх "горку", крутизна которой падает, если точка (x_1, x_2) перемещается на "северо-восток" по плоскости Ox_1x_2.

Пример 1.3. Линейная ПФ (ЛПФ) имеет вид: $y = a_0 + a_1x_1 + a_2x_2$ (двухфакторная) и $y = a_0 + a_1x_1 + \ldots + a_nx_n$ (многофакторная). ЛПФ принадлежит к классу так называемых *аддитивных* ПФ (АПФ). Переход от мультипликативной ПФ к аддитивной осуществляется с помощью операции логарифмирования. Для двухфакторной мультипликативной ПФ

$$y = a_0 x_1^{a_1} x_2^{a_2}$$

этот переход имеет вид: $\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2$.

Полагая $\ln y = w$, $\ln x_1 = v_1$ и $\ln x_2 = v_2$, получаем аддитивную ПФ $w = \ln a_0 + a_1 v_1 + a_2 v_2$.

Выполняя обратный переход, из аддитивной ПФ получим мультипликативную ПФ.

Если сумма показателей степени в ПФ Кобба-Дугласа $y = a_0K^{a_1}L^{a_2}$ равна единице ($a_1 + a_2 = 1$), то ее можно записать в несколько другой форме:

$$\frac{Y}{L}=\frac{a_0K^{a_1}L^{a_2}}{L}=\frac{a_0K^{a_1}}{L^{1-a_2}}=\frac{a_0K^{a_1}}{L^{a_1}}=a_0\left[\frac{K}{L}\right]^{a_1}, \text{ т.е. } \frac{Y}{L}=a_0\left[\frac{K}{L}\right]^{a_1}$$

Дроби $\frac{Y}{L} = z$ и $\frac{K}{L} = k$ называются соответственно *производительностью труда* и *капиталовооруженностью труда*. Используя новые символы, получим

$$z = a_0k^{a_1},$$

т.е. из двухфакторной ПФКД получим формально однофакторную ПФКД. В связи с тем, что $0 < a_1 < 1$, из последней формулы следует, что производительность труда z растет медленнее его капиталовооруженности. Однако этот вывод справедлив для случая статической ПФКД в рамках существующих технологии и ресурсов.

Отметим здесь, что дробь $\frac{Y}{K}$ называется *производительностью капитала* или *капиталоотдачей,* обратные дроби $\frac{K}{Y}$ и $\frac{L}{Y}$ называются соответственно *капиталоемкостью* и *трудоемкостью* выпуска.

ПФ называется *динамической,* если:

1) время t фигурирует в качестве самостоятельной переменной величины (как бы самостоятельного фактора производства), влияющего на объем выпускаемой продукции;

2) параметры ПФ и ее характеристика f зависят от времени t.

Отметим, что если параметры ПФ *оценивались* по данным *временных рядов* (объемов ресурсов и выпуска) продолжительностью T_0 лет (т.е. *базовый промежуток* для оценки параметров имеет продолжительность T_0 лет), то экстраполяционные расчеты по такой ПФ следует проводить не более чем на $\frac{T_0}{3}$ лет вперед (т.е. *промежуток экстраполяции* должен иметь продолжительность не более чем $\frac{T_0}{3}$ лет).

При построении ПФ *научно-технический прогресс* (НТП) может быть учтен с помощью введения *множителя* НТП e^{pt}, где параметр (число) p ($p > 0$) характеризует *темп прироста* выпуска под влиянием НТП:

$$y(t) = e^{pt}f(x_1(t), x_2(t)) \ (t = 0, 1, \ldots, T).$$

Эта ПФ - простейший пример *динамической* ПФ; она включает *нейтральный*, то есть не материализованный в одном из факторов, технический прогресс. В более сложных случаях технический прогресс может воздействовать непосредственно на производительность труда или капиталоотдачу: $Y(t) = f(A(t)\cdot L(t), K(t))$ или $Y(t) = f(A(t)$ $K(t), L(t))$. Он называется, соответственно, трудосберегающим или капиталосберегающим НТП.

Пример 1.4. Приведем вариант ПФКД с учетом НТП $y(t) = a_0 e^{pt}\cdot x_1(t)^{a_1} x_2(t)^{a_2}$.

Выделение существенных видов ресурсов (факторов производства) и *выбор* аналитической формы функции $f(x_1, x_2)$ называется *спецификацией* ПФ $y = f(x_1, x_2)$.

Преобразование реальных и экспертных данных в модельную информацию, т.е. *расчет* численных значений параметров ПФ $y = f(x_1, x_2)$ на базе статистических данных с помощью регрессионного и корреляционного анализа, называется *параметризацией* ПФ $y = f(x_1, x_2)$.

Проверка истинности (адекватности) ПФ называется ее *верификацией.*

Выбор аналитической формы ПФ $y = f(x_1, x_2)$ (т.е. спецификация) диктуется прежде всего *теоретическими* соображениями, которые должны явно (или даже неявно) учитывать особенности взаимосвязей между конкретными ресурсами (в случае микроэкономического уровня) или экономических закономерностей (в случае макроэкономического уровня), особенности реальных или экспертных данных, преобразуемых в параметры ПФ (т.е. особенности параметризации). На спецификацию и параметризацию в процессе совершенствования ПФ оказывают влияние результаты верификации ПФ. Отметим здесь, что оценка параметров ПФ обычно проводится с помощью *метода наименьших квадратов*.

Пример 1.5. Приведем в качестве иллюстрации значения параметров a_1 и a_2 макроэкономической ПФ Кобба-Дугласа $Y = a_0 K^{a_1} L^{a_2}$ для экономики США, рассчитанные разными авторами для разных базовых временных промежутков (априори не предполагалось, что обязательно $a_1 + a_2 = 1$) (см.: Терехов Л.Л. Производственные функции. М.: Статистика, 1974, с. 113):

Базовые временные промежутки (или годы)	Параметры			Авторы, проводившие исследования
	a1	*a2*	*a1 + a2*	
1899-1922 гг.	0,25	0,75	1,00	Дуглас
1904 г.	0,31	0,65	0.96	Дуглас
1914 г.	0,36	0,61	0,97	Дуглас
1919 г.	0,25	0,76	1,01	Дуглас
1869-1948 гг.	0,70	0,25	0,95	Валаванис
1900-1953 гг.	0,16	0,84	1,00	Клейн
1909-1949 гг.	0,35	0,65	1,00	Солоу
1921-1941 гг.	0,34	2,13	2,47	Тинтнер
1934-1959 гг.	0,41	0,91	1,32	Михалевский
1934-1956 гг.	0,26	0,74	1,00	Михалевский

Параметры разными авторами рассчитывались по разным методикам, поэтому пестрота картины не является неожиданной. Обращает на себя внимание, что у всех авторов наблюдается значительное превышение параметра a_2 относительно параметра a_1. Также почти у всех авторов сумма $a_1 + a_2$ оказалась близкой к единице.

На основании данных по экономике СССР (динамика национального дохода, численности занятых в материальном производстве и объемов основных фондов), опубликованных за 1960-1985 гг., были рассчитаны параметры МАПФКД без учета НТП и с учетом НТП. Без учета НТП ПФКД имеет вид $Y = 1{,}022\, K^{0.5382} L^{0{,}4618}$ (коэффициент детерминации $R^2 = 0{,}9969$, статистика Дарбина-Уотсона $DW = 0{,}81$; упомянутые здесь термины математической статистики будут проанализированы в главах 15, 16). При подстановке фактических значений K и L за 1986 г. ошибка прогноза с помощью выписанной ПФКД составила 3%, что свидетельствует о том, что точность прогноза на основе рассматриваемой ПФКД относительно невелика. С учетом НТП ПФКД имеет вид $Y = 1{,}038\, e^{0.0294t} K^{0{,}9749} L^{0{,}2399}$ (коэффициент детерминации $R^2 = 0{,}9982$, статистика Дарбина-Уотсона $DW = 1{,}63$) (см.: Гранберг А.Г. Моделирование социалистической экономики. М.: Экономика, 1988).

10.2. Формальные свойства производственных функций

Производственная функция $f(x_1, x_2)$ как формальная конструкция определена в *неотрицательном ортанте* двумерной плоскости. т.е. определена при $x_1 \geq 0$, $x_2 \geq 0$. ПФ должна удовлетворять ряду (для каждой конкретной ПФ - своему) *свойств:*

1. $f(0,0)=0$;

1". $f(0,x_2)=f(x_1,0)=0$;

2. $x(1)\geq x(0)$ $(x(1) \neq x(0)) \Rightarrow f(x(1))>f(x(0))$ $(x(k)=(x_1(k),x_2(k)$, $k=0,1))$;

2". $x > 0 \Rightarrow \dfrac{\partial f(x)}{\partial x_i} > 0$ $(i=1,2)$, $x=(x_1,x_2)$;

3. $x > 0 \Rightarrow \dfrac{\partial^2 f(x)}{\partial x_i^2} \leq 0$ $(i=1,2)$, $x=(x_1,x_2)$;

3". $x > 0 \Rightarrow \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} \geq 0$ $x=(x_1,x_2)$;

4. $f(tx_1,tx_2)=t^p f(x_1,x_2)$.

Свойство 1 означает, что без ресурсов нет выпуска. Свойство 1" означает, что при отсутствии хотя бы одного из ресурсов нет выпуска.

Свойство 2 означает, что с ростом затрат хотя бы одного ресурса объем выпуска растет. Свойство 2" (первая частная производная ПФ $\left(\dfrac{\partial f(x)}{\partial x_i}\right)$ положительна) означает, что с ростом затрат одного ресурса при неизменном количестве другого ресурса объем выпуска растет. Упорядоченная пара (x_1,x_2) чисел x_1 и x_2 для краткости здесь и далее обозначается символом x, т.е. $x=(x_1,x_2)$.

Свойство 3 (вторая частная производная ПФ $\left(\dfrac{\partial^2 f(x)}{\partial x_i^2}\right)$ неположительна) означает, что с ростом затрат одного (i-го) ресурса при неизменном количестве другого ресурса величина прироста выпуска на каждую дополнительную единицу i-го ресурса не растет (*закон убывающей эффективности*). Свойство 3" означает, что при росте одного ресурса предельная эффективность другого ресурса возрастает. Если выполнены условия 3-3", то график $Г$ ПФ есть поверхность, расположенная в неотрицательном ортанте $x_1 \geq 0$, $x_2 \geq 0$, $y \geq 0$ трехмерного пространства Ox_1x_2y и выпуклая вверх. Вообще геометрический образ ПФ должен прежде всего ассоциироваться с выпуклой горкой, крутизна которой убывает, если точка (x_1, x_2) уходит в плоскости Ox_1x_2 на “северо-восток”.

Свойство 4 означает, что ПФ является *однородной функцией* (ОФ) *степени* $p > 0$. При $p > 1$ с ростом масштаба производства в t раз (число $t > 1$), т.е. с переходом от вектора x к вектору tx, объем выпуска возрастает в t^p ($> t$) раз, т.е. имеем *рост эффективности* производства от *роста масштаба* производства. При $p < 1$ имеем *падение эффективности* производства от *роста масштаба* производства. При $p = 1$ имеем *постоянную эффективность* производства при *росте* его *масштаба* (или имеем независимость удельного выпуска от масштаба производства - в английской терминологии constant returns to scale).

Для ПФКД $y = a_0 x_1^{a_1} x_2^{a_2}$ ($a_1 + a_2 = 1$) свойства 1-4 выполняются.

Для ЛПФ $y=a_0+a_1x_1+a_2x_2$ ($a_0>0$, $a_1>0$, $a_2>0$) свойства 1 и 1' (при $a_0=0$) и свойство 4 не выполняются.

Множество (линия) l_q уровня $q = f(x_1, x_2)$ ($0 < q$ - действительное число) ПФ $y = f(x_1, x_2)$ называется *изоквантой* ПФ. Иными словами, линия уровня q - это множество точек, в котором ПФ постоянна и равна q.

Различные наборы (v_1, v_2) и (w_1, w_2) затрачиваемых (используемых) ресурсов, принадлежащие одной и той же изокванте l_q (т.е. $q = f(v_1, v_2) = f(w_1, w_2)$), дают один и тот же объем выпуска q. Изокванта есть линия, расположенная в неотрицательном ортанте $\{(x_1, x_2) | x_1 \geq 0, x_2 \geq 0\}$ двумерной плоскости Ox_1x_2.

Пример 2.1. На рис. 10.4 даны эскизы изоквант l_{q1} и l_{q2} ПФКД. Отметим, что изокванта l_{q2}, расположенная "северо-восточнее" изокванты l_{q1}, соответствует большему объему выпуска (т.е. $q_2 > q_1$). Если объем используемого основного капитала неограниченно растет (т.е. $x_1 = K \to +\infty$), то, как видно на рис. 10.4, затраты труда неограниченно убывают (т.е. $x_2 = L \to +0$). Аналогично, как видно на рис. 10.4, если $x_2 = L \to +\infty$, то $x_1 = K \to +0$. На рис. 10.5 даны эскизы изоквант l_{q1} и l_{q2} ($q_2 > q_1$) ЛПФ.

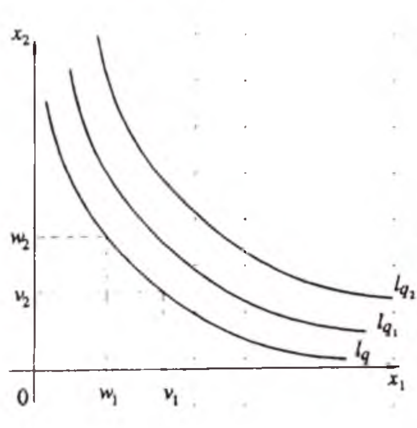


Рис. 10.4

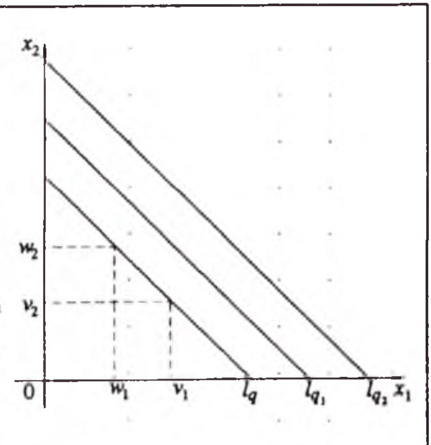


Рис. 10.5

При n=2 для любой ПФ, для которой справедливы все (или часть) свойств 1-4, изокванта (если она не является прямой) есть линия (не обязательно гладкая), которая *выпукла* к точке О, т.е. линия, которая похожа на изокванту l_q рис. 4. Если график $Г$ ПФ похож на выпуклую горку, то естественно, что ее изокванты есть линии, выпуклые к точке О.

10.3. Предельные (маржинальные) и средние значения производственной функции

Пусть $y = f(x) = f(x_1, x_2)$ - ПФ. Дробь

$$\frac{f(x)}{x_i} \quad (i = 1,2)$$

называется *средней производительностью* i-го ресурса (***фактора*** производства) (СПФ) или средним выпуском по i-му ресурсу (фактору производства). Символика: $A_i = \frac{f(x)}{x_i}$.

Напомним, что в случае двухфакторной ПФКД $Y = a_0K^{a_1}L^{a_2}$ для средних производительностей $\frac{Y}{K}$ и $\frac{Y}{L}$ основного капитала и труда были использованы соответственно термины капиталоотдача и производительность труда. Эти термины используют и применительно к любым двухфакторным ПФ, у которых $x_1=K$ и $x_2=L$.

Пусть $y = f(x) = f(x_1, x_2)$ - ПФ. Ее первая частная производная

$$\frac{\partial f(x)}{\partial x_i} \quad (i = 1,2)$$

называется *предельной (маржинальной) производительностью* i-го ресурса (***фактора*** производства) (ППФ) или предельным выпуском по i-му ресурсу (фактору производства). Символика: $M_i = \frac{\partial f(x)}{\partial x_i}$.

Обозначим символами Δx_i и $\Delta_i(f(x))$ ($\Delta_1 f(x_1,x_2)=f(x_1+\Delta x_1,x_2)-f(x_1,x_2)$; $\Delta_2 f(x_1,x_2)=f(x_1,x_2+\Delta x_2)-f(x_1,x_2)$) соответственно, приращение переменной x_i и соответствующее ей частное приращение ПФ $f(x)$. При малых Δx_i имеем приближенное равенство

$$\frac{\partial f(x)}{\partial x_i} \approx \frac{\Delta_i f(x)}{\Delta x_i} \quad (i = 1,2).$$

Следовательно, ППФ (приближенно) показывает, на *сколько* единиц *увеличится* объем выпуска y, если объем затрат x_i i-го ресурса *вырастает* на *одну* (достаточно малую) единицу при неизменных *объемах* другого затрачиваемого ресурса. Здесь *предельную* величину $\frac{\partial f(x)}{\partial x_i}$ (т.е. ППФ) целесообразно интерпретировать, используя близкое к ней отношение малых *конечных* величин, т.е. $\Delta f(x)$ и Δx_i. Отмеченное обстоятельство является ключевым для понимания экономического смысла ППФ $\frac{\partial f(x)}{\partial x_i}$. С другими предельными величинами следует поступать аналогичным образом.

Задача 3.1. Для ПФКД $y = a_0 x_1^{a_1} x_2^{a_2}$ найти в явном виде A_1, A_2, M_1 и M_2.

Решение задачи. Имеем:

$$A_1 = \frac{y}{x_1} = \frac{f(x)}{x_1} = a_0 x_1^{a_1-1} x_2^{a_2}; \; A_2 = \frac{y}{x_2} = \frac{f(x)}{x_2} = a_0 x_1^{a_1} x_2^{a_2-1};$$

$$M_1 = \frac{\partial f(x)}{\partial x_1} = a_1 A_1; \; M_2 = \frac{\partial f(x)}{\partial x_2} = a_2 A_2;$$

$$\frac{M_1}{A_1} = a_1 \le 1 \Rightarrow M_1 \le A_1; \; \frac{M_2}{A_2} = a_2 \le 1 \Rightarrow M_2 \le A_2.$$

Для ПФ $y = f(x)$ (не только для ПФКД) неравенства

$$M_i \le A_i \; (i = 1, 2)$$

(т.е. *предельная производительность* i-го ресурса *не больше средней* производительности этого ресурса) обычно выполняются.

Задача 3.2. Для ЛПФ $y=a_0+a_1x_1+a_2x_2$ ($a_0>0$, $a_1>0$, $a_2>0$) найти в явном виде A_1, A_2, M_1 и M_2.

Решение задачи. Имеем:

$$A_1 = \frac{y}{x_1} = \frac{f(x)}{x_1} = \frac{a_0}{x_1} + a_1 + a_2\frac{x_2}{x_1}; \; A_2 = \frac{y}{x_2} = \frac{f(x)}{x_2} = \frac{a_0}{x_2} + a_1\frac{x_1}{x_2} + a_2,$$

$$M_1 = \frac{\partial f(x)}{\partial x_1} = a_1, \; M_2 = \frac{\partial f(x)}{\partial x_2} = a_2,$$

$$\frac{M_1}{A_1} \le 1 \Rightarrow M_1 \le A_1; \; \frac{M_2}{A_2} \le 1 \Rightarrow M_2 \le A_2.$$

Пусть $y = f(x)$ - ПФ, $x = (x_1, x_2)$. Отношение предельной производительности M_i i-го ресурса к его средней производительности A_i

называется (частной) *эластичностью выпуска* по i-му ресурсу (по *фактору* производства) (ЭВФ). Символика:

$$E_i = \frac{M_i}{A_i} = \frac{x_i}{f(x)}\frac{\partial f(x)}{\partial x_i}\ (i = 1,2).$$

Сумма $E_1 + E_2 = E_x$ называется *эластичностью производства*.

Поскольку при малом приращении Δx_i имеем приближенное равенство

$$E_i = \left[\frac{\partial f(x)}{\partial x_i}\right] \Big/ \left[\frac{f(x)}{x_i}\right] \approx \left[\frac{\Delta_i f(x)}{f(x)}\right] \Big/ \left[\frac{\Delta x_i}{x_i}\right]$$

(крайнее правое выражение есть отношение двух относительных величин $\frac{\Delta_i f(x)}{f(x)}$ и $\frac{\Delta x_i}{x_i}$), постольку E_i (приближенно) показывает, *на сколько процентов* увеличится выпуск y, если затраты i-го ресурса увеличатся *на один* процент при неизменных объемах другого ресурса. Пояснение выражения E_i, содержащего предельную величину $\frac{\partial f(x)}{\partial x_i}$, с помощью выражения, содержащего конечное приближение $\frac{\Delta_i f(x)}{f(x)}$ этой предельной величины, является *ключевым* в понимании экономической сути частной эластичности выпуска по i-му ресурсу.

Задача 3.3. Выписать в явном виде для ПФКД выражения для E_1, E_2 и E_x.

Решение задачи. Имеем:

$E_1 = a_1$, $E_2 = a_2$,

$E_x = E_1 + E_2 = a_1 + a_2$.

Задача 3.4. Для ЛПФ $y=a_1x_1+a_2x_2$ ($a_0=0$) выписать в явном виде выражения для E_1, E_2 и E_x.

Решение задачи. Имеем:

$$E_1 = \frac{x_1}{f(x)}\frac{\partial f(x)}{\partial x_1} = \frac{a_1x_1}{a_1x_1 + a_2x_2};\quad E_2 = \frac{x_2}{f(x)}\frac{\partial f(x)}{\partial x_2} = \frac{a_2x_2}{a_1x_1 + a_2x_2};$$

$E_x = E_1 + E_2 = 1$.

Пусть $y = f(x)$ - ПФ, $x = (x_1,x_2)$. *Предельной нормой замены* (замещения) i-го ресурса (*фактора* производства) j-м (аббревиатура: ПНЗФ и символика: R_{ij}) называется выражение

$$R_{ij} = -\frac{dx_j}{dx_i}\ (i, j = 1,2) \tag{3}$$

при постоянной y.

Обратим внимание на то, что i - номер *заменяемого* ресурса, j - номер *замещающего* ресурса. Используется также термин: предельная технологическая норма замены (замещения) i-ого ресурса (фактора производства) j-м ресурсом (фактором производства). Приведем более краткий (но менее точный) термин: (предельная) норма замены (замещения) ресурсов.

Пусть выпуск y является постоянным (т.е. все наборы затрачиваемых ресурсов расположены на одной изокванте), тогда первый полный дифференциал dy ПФ $y = f(x)$ тождественно равен нулю:

$$0 = dy = \frac{\partial f(x)}{\partial x_1}dx_1 + \frac{\partial f(x)}{\partial x_2}dx_2$$

(здесь dx_1, dx_2 - дифференциалы переменных x_1, x_2), откуда, выражая первый дифференциал dx_j, получим ($i \neq j$):

$$dx_j = -\frac{\frac{\partial f(x)}{\partial x_i}}{\frac{\partial f(x)}{\partial x_j}}dx_i \quad (i, j = 1,2), \tag{4}$$

откуда, поделив на dx_i, получим

$$\frac{dx_j}{dx_i} = -\frac{\frac{\partial f(x)}{\partial x_i}}{\frac{\partial f(x)}{\partial x_j}} \quad (i, j = 1,2). \tag{5}$$

На основании (3), (4) и (5) имеем:

$$R_{ij} = -\frac{dx_j}{dx_i} = \frac{\frac{\partial f(x)}{\partial x_i}}{\frac{\partial f(x)}{\partial x_j}} > 0 \quad (i \neq j,\ i = 1,2). \tag{6}$$

Отметим, что строгий вывод формулы (6) опирается в действительности на теорему о неявной функции, формулировка которой в настоящем пособии не приводится.

Непосредственно проверяется, что для двухфакторной ПФ справедливо равенство

$$R_{12} = \frac{E_1}{E_2}\frac{x_2}{x_1},$$

т.е. (предельная) норма замены первого ресурса вторым равна отношению эластичностей выпуска по первому и второму ресурсам, умноженному на отношение объема второго ресурса к объему первого ресурса. Если $x_1 = K$, $x_2 = L$, то отношение $\frac{x_1}{x_2} = \frac{K}{L}$ называется капиталовооруженностью труда. В этом случае (предельная) норма замены основного капитала трудом равна отношению эластичностей выпуска по основному капиталу и труду, поделенному на капиталовооруженность труда.

Пусть ПФ - двухфакторная. При постоянном выпуске y и малых приращениях Δx_1 и Δx_2 имеем приближенное равенство

$$R_{12} = -\frac{dx_2}{dx_1} \approx -\frac{\Delta x_2}{\Delta x_1}. \tag{7}$$

На основании (7) (предельная) норма замены ресурсов R_{12} (приближенно) показывает, *на сколько единиц увеличатся* затраты второго ресурса (при неизменном выпуске $y = a$), если затраты первого ресурса *уменьшатся* на *одну* (малую) *единицу*. См. рис. 10.6, на котором видно, что чем круче касательная к изокванте $l(q)$ в точке (x_1, x_2), тем больше выражение $-\frac{dx_2}{dx_1}$ и, следовательно, норма замены R_{12} первого ресурса вторым.

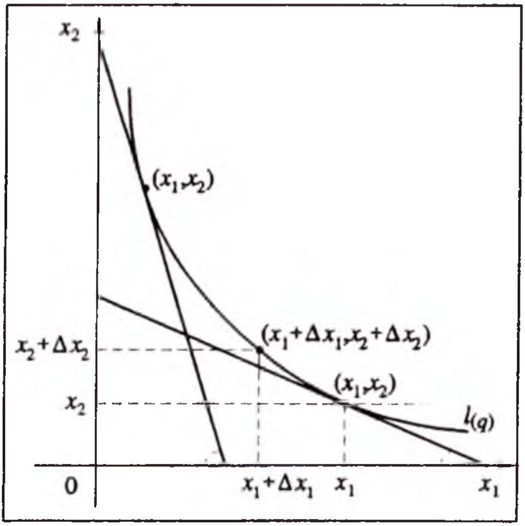


Рис. 10.6

Задача 3.5. Для ПФКД $y = a_0x_1^{a_1}x_2^{a_2}$ выписать в явном виде выражения R_{12} и R_{21}.

Решение задачи. Имеем:

$$R_{12} = \left|\frac{\partial y}{\partial x_1}\right| \Big/ \left|\frac{\partial y}{\partial x_2}\right| = \frac{a_1}{a_2}\frac{x_2}{x_1};\ R_{21} = \left|\frac{\partial y}{\partial x_2}\right| \Big/ \left|\frac{\partial y}{\partial x_1}\right| = \frac{a_2}{a_1}\frac{x_1}{x_2}.$$

Задача 3.6. Для ПФ $y=a_0+a_1x_1+a_2x_2$ выписать в явном виде выражения R_{12} и R_{21}.

Решение задачи. Имеем:

$$R_{12} = \left|\frac{\partial y}{\partial x_1}\right| \Big/ \left|\frac{\partial y}{\partial x_2}\right| = \frac{a_1}{a_2};\ R_{21} = \left|\frac{\partial y}{\partial x_2}\right| \Big/ \left|\frac{\partial y}{\partial x_1}\right| = \frac{a_2}{a_1}.$$

10.4. Производственные функции в темповой записи

Наряду со связями объемных показателей выпуска и затрат ресурсов могут быть рассмотрены связи между темпами прироста этих показателей. Будем здесь говорить о макроэкономических производственных функциях, связывающих величину совокупного продукта (дохода) Y с затратами капитала K и труда L, но все это легко обобщается на любые другие производственные функции. Обозначим темпы прироста величин Y, K и L малыми буквами y, k и l соответственно. Это могут быть дискретные темпы прироста $y_t = \frac{Y_t-Y_{t-1}}{Y_{t-1}}$, $k_t = \frac{K_t-K_{t-1}}{K_{t-1}}$, $l_t = \frac{L_t-L_{t-1}}{L_{t-1}}$ или непрерывные темпы прироста $y_t = \frac{Y_t'}{Y_t}$, $k_t = \frac{K_t'}{K_t}$, $l_t = \frac{L_t'}{L_t}$. Итак, ПФ в темповой записи имеет вид: $y = f(k,l)$.

Теперь рассмотрим связь ПФ Кобба-Дугласа в объемной и темповой записи. Пусть величины K и L являются непрерывными дифференцируемыми функциями времени (K_t и L_t). В таком случае они представляют не объемы использованных ресурсов за определенный период времени, а "интенсивности" их использования в каждый момент времени. От функции $Y_t=AK_t^{\alpha}L_t^{\beta}e^{\gamma t}$ можно после ее логарифмирования взять полный дифференциал:

$$d\ln Y_t=\alpha\cdot d\ln K_t+\beta\cdot d\ln L_t+\gamma dt \text{ или}$$

$$\frac{dY_t}{Y_t} = \alpha\cdot\frac{dK_t}{K_t} + \beta\cdot\frac{dL_t}{L_t} + \gamma dt \Rightarrow$$

$$\frac{Y_t'dt}{Y_t} = \alpha\cdot\frac{K_t'dt}{K_t} + \beta\cdot\frac{L_t'dt}{L_t} + \gamma dt.$$

и после деления обеих частей на dt получаем

$$\frac{Y_t'}{Y_t} = \alpha\cdot\frac{K_t'}{K_t} + \beta\cdot\frac{L_t'}{L_t} + \gamma,$$

Здесь $y_t = \frac{Y_t'}{Y_t}$, $k_t = \frac{K_t'}{K_t}$, $l_t = \frac{L_t'}{L_t}$ - непрерывные темпы прироста выпуска, капитала и труда. Таким образом, ПФ Кобба-Дугласа в объемных показателях соответствует линейная зависимость темпов прироста:

$$y_t = \alpha k_t + \beta l_t + \gamma.$$

Эта зависимость называется ПФ Кобба-Дугласа в темповой записи.

Если заменить дифференциалы dY_t, dK_t, dL_t (главные линейные части приращений) на сами приращения ΔY_t, ΔK_t, ΔL_t, то получим приближенную формулу $y_t = \alpha k_t + \beta l_t + \gamma$, где y_t, k_t, l_t - дискретные темпы прироста. Таким образом, и в дискретном случае функции Кобба-Дугласа в объемных показателях соответствует линейная формула связи темпов прироста y_t, k_t и l_t. Однако при ее анализе и оценивании надо иметь в виду следующее. Формулы $Y_t = AK_t^{\alpha}L_t^{\beta}e^{\gamma t}$ и $y_t = \alpha k_t + \beta l_t + \gamma$ эквивалентны при непрерывном рассмотрении времени. В то же время статистические данные, по которым оцениваются ПФ, всегда дискретны; обычно это погодовые данные. В этих условиях приведенные формулы зависимостей для объемов и темпов прироста - это разные ПФ. Иногда оценки параметров α, β и γ, полученные для объемной ПФ Кобба-Дугласа, переносят на темповую формулу, и наоборот. Так делать некорректно; каждая из этих формул должна быть оценена в отдельности. Даже если они оценены по одним и тем же статистическим данным (то есть по объемам и темпам, соответствующим друг другу), результаты такой оценки могут быть совершенно различными. Одна из формул, например, может не дать статистически значимой оценки, в то время как по другой получается вполне приемлемый результат.

Из приведенных соображений вытекает, что показатель γ - свободный член ПФ Кобба-Дугласа в темповой записи - это темп нейтрального технического прогресса. Это та часть темпа прироста выпуска, которая не связана с приростом затрат капитала и труда, а отражает интенсификацию производства на макроуровне.

Пусть, например, оценена следующая формула ПФ в темповой записи:

$$y_t = 0{,}3 \cdot k_t + 0{,}6 \cdot l_t + 1{,}5.$$

Пусть при этом средний темп прироста затрат труда l_t составил 1%, средний темп прироста используемого капитала k_t - 6%, а средний темп прироста выпуска y_t - 3,9%. Вклад в эту цифру экстенсивных факторов - прироста затрат капитала и труда - составил соответственно $0{,}3 \cdot 6 = 1{,}8$ (%) и $0{,}6 \cdot 1 = 0{,}6$ (%). Вклад интенсивных факторов (технического прогресса) составляет 1,5 процентных пункта, или $\frac{1{,}5}{3{,}9} \cdot 100\% \approx 38{,}5\%$.

10.5. Эластичность замещения факторов. Производственная функция CES

Обобщение ПФ Кобба-Дугласа может вестись в различных направлениях. Наиболее известным обобщением является функция CES, или ПЭЗ, - функция с постоянной эластичностью замещения (constant elasticity of substitution). Эластичность замещения σ - это мера "кривизны" изоквант (линий уровня) ПФ. Точнее, "кривизну" измеряет величина $\frac{1}{\sigma}$. Эластичность замещения труда капиталом $\sigma_{LK} = d\ln\left(\frac{K}{L}\right) \Big/ d\ln\left(\frac{Y_L'}{Y_K'}\right)$ показывает, на сколько процентов изменится капиталовооруженность $\left(\frac{K}{L}\right)$ при изменении предельной нормы замены труда капиталом $\left(MRS_{KL} = -\frac{dK}{dL} = \frac{Y_L'}{Y_K'}\right)$ на 1%. Если изобразить одну из изоквант (линий уровня, т.е. $Y = \text{const}$) ПФ на плоскости KL (см. рис. 10.7), обозначив ее цифрой 1, то предельная норма замены в точке A - это тангенс угла наклона этой изокванты (то есть $\operatorname{tg}\alpha$).

При перемещении из точки A в точку B по изокванте наклон касательной меняется, меняется и соотношение $\left(\frac{K}{L}\right)$. Это соотношение постоянно вдоль каждой прямой, проходящей через начало координат (например, прямых 2 и 3). Величина $\frac{1}{\sigma}$ показывает относительное изменение тангенса угла наклона линии уровня в расчете на

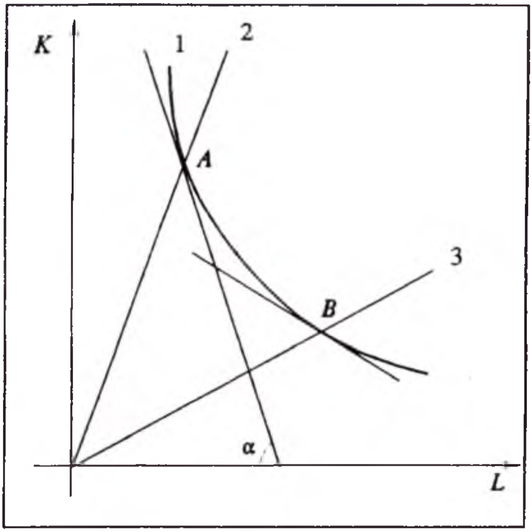


Рис. 10.7

единицу изменения отношения $\left|\frac{K}{L}\right|$. Очевидно, чем сильнее меняется наклон линии уровня при переходе, скажем из точки A в точку B (с прямой 2 на прямую 3), тем больше "кривизна" линии уровня. На рисунках (10.8а) - (10.8г) изображены линии уровня, соответственно, ПФ $Y = aK + bL + c$ (линейной), ПФ Кобба-Дугласа (б), ПФ с бесконечной эластичностью замещения $Y = \min(aK, bL)$ (функция Леонтьева) и производственной функции CES (функции с постоянной эластичностью замещения, вопросы оценки и анализа которой мы рассмотрим ниже).

Линейная ПФ имеет нулевую "кривизну" и, соответственно, бесконечную эластичность замещения у. Функция Кобба-Дугласа имеет эластичность замещения, равную единице. Функция Леонтьева имеет нулевую эластичность замещения: ресурсы в ней должны использоваться в заданной пропорции и не могут замещать друг друга. В реальной экономике степень взаимозаменяемости ресурсов может быть различной, соответственно различной (а не только нулевой, бесконечной или единичной) может быть и эластичность замещения. Это ставит задачу оценки более общих формул ПФ, в частности ПФ с постоянной, но произвольной эластичностью замещения. Такая функция (функция CES) описывается формулой

$$Y = A\,(uK^{-\rho} + (1-u)L^{-\rho})^{-n/\rho}.$$

Рисунок 10.8. Семейства линий уровня для различных ПФ

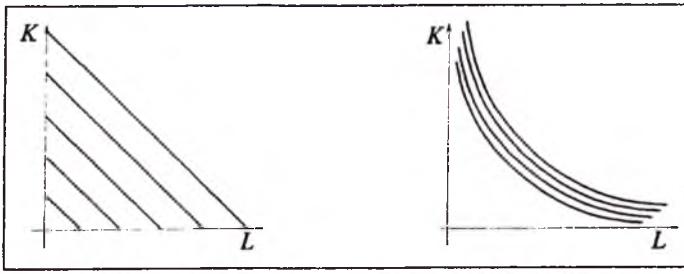


Рисунок 10.8а. Функция $Y = aK + bL + c$ (линейная)

Рисунок 10.8б. Функция $Y = A\ K^{\alpha} L^{\beta}$ (Кобба-Дугласа)

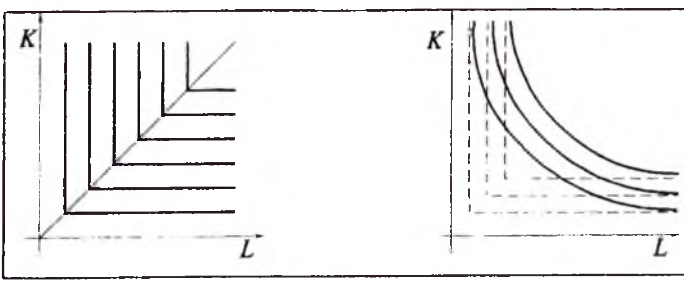


Рисунок 10.8в. Функция $Y = \min(aK, bL)$ (Леонтьева)

Рисунок 10.8г. Функция CES

Здесь $\rho \geq -1$; $n>0$ - степень однородности; $A>0$; $0<u<1$. Эластичность замещения для такой функции равна $\frac{1}{1+\rho}$. Если $\rho=-1$, то получаем функцию с линейными изоквантами (в частности, линейную), при $\rho \to 0$ в пределе получаем ПФ Кобба-Дугласа с $\sigma=1$, при $\rho \to \infty$ - ПФ Леонтьева.

В качестве примера оценки ПФ CES приведем полученные различными авторами результаты для экономики СССР. Такие оценки делались за различные периоды времени в промежутке 1950-1985 гг. Э.Б.Ершовым, Ю.В.Яременко и А.С.Смышляевым, М.Вейтцманом, А.Г.Гранбергом, Н.Б.Баркаловым и другими. Исходные спецификации различаются предпосылками о степени однородности n (в большинстве случаев изначально считалось, что $n = 1$, но были и оценки с произвольным n) и наличием множителя e^{rt}, характеризующего нейтральный технический прогресс (такой множитель может добавляться не только к ПФ Кобба-Дугласа, но и к CES или какой-либо другой функции). Например, А.Г.Гранберг приводит следующие оценки за 1960-1985 гг.:

$$Y = 1{,}002\,(0{,}6412\,K^{-0{,}81}+0{,}3588\,L^{-0{,}81})^{-1/0{,}81}\ ;$$
$$R^2 = 0{,}9984;\ DW = 1{,}58$$

(линейно-однородная функция CES без учета технического прогресса);

$$Y = 0{,}966\,(0{,}4074\,K^{-3{,}03}+0{,}5926\,L^{-3{,}03})^{-1/3{,}03}\ e^{0{,}0252t};$$
$$R^2 = 0{,}9982;\ DW = 1{,}76$$

(линейно-однородная функция CES с учетом технического прогресса).

С точки зрения статистик R^2 и DW, обе зависимости получились значимыми. В то же время оценки показателя эластичности замещения $\sigma = \dfrac{1}{1+\rho}$ в них различны: в первом случае это 0,55, во втором - 0,25. Другими авторами оценки эластичности замещения у для экономики СССР также получены меньшими единицы: 0,2 (Ю.В.Яременко и др.), 0,4 (М.Вейтцман), 0,37-0,43 за разные периоды (Н.Б.Баркалов). В целом можно сказать, что оценка эластичности замещения очень зависела от конкретной спецификации, но в большинстве случаев составляла около 0,4. Во всяком случае, она заведомо была для экономики СССР меньше единицы, что говорит о невысокой степени взаимозаменяемости труда и капитала. Эта взаимозаменяемость была гораздо ниже, чем это предполагается в функции Кобба-Дугласа, в которой эластичность замещения у априори считается равной единице. Ошибочность исходной гипотезы о степени взаимозаменяемости факторов может служить причиной недостаточной статистической значимости оценок ПФ Кобба-Дугласа.

Вопросы к главе 10

1. В чем суть закона убывающей эффективности?
2. Что в статической производственной функции не зависит от времени t, а что может зависеть от времени t?
3. Как определяется (средняя) производительность труда и капиталовооруженность (фондовооруженность) труда? Какие возможны варианты взаимосвязи между ними в случае производственной функции Кобба-Дугласа?
4. Назовите основные свойства, которыми должна обладать производственная функция. Приведите примеры производственных функций, которые отдельными свойствами не обладают. Приведите примеры производственных функций, которые обладают всеми основными свойствами.
5. Что такое изокванта? В чем ее экономический смысл?
6. Как определяется (средняя) производительность капитала (капиталоотдача)?
7. Как определяется (предельная) производительность капитала и (предельная) производительность труда?
8. Какая существует связь между средней и предельной производительностью капитала (труда) в общем случае и в случае производственной функции Кобба-Дугласа?
9. Сформулируйте определение (частной) эластичности выпуска по i-му ресурсу (i-му фактору производства) ($i = 1, 2$) и определение эластичности производства.
10. Дайте содержательную интерпретацию (частной) эластичности выпуска по i-му ресурсу.
11. Сформулируйте определение (предельной) нормы замены одного ресурса другим. Дайте содержательную интерпретацию этому понятию.
12. Как меняется (предельная) норма замены одного ресурса другим при движении по изокванте? Дайте содержательную интерпретацию характеру изменения предельной нормы замены.
13. Дайте определение и поясните смысл ПФ в темповой записи.
14. Как связаны ПФ Кобба-Дугласа в объемной и темповой записи?
15. Как описывается технический прогресс в ПФ в объемной и темповой записи? Как оценить долю вклада интенсивных факторов в темпы экономического роста?
16. Дайте определение и графическую интерпретацию эластичности замещения факторов.
17. Поясните смысл ПФ CES. Каковы ее свойства и основные характеристики?

ГЛАВА 11
ЗАДАЧИ ОПТИМИЗАЦИИ ПРОИЗВОДСТВА

11.1. Основные понятия

Доходом (выручкой) R фирмы в определенном временном периоде (например, в определённом году) называется произведение $p_0 y$ общего объема y выпускаемой фирмой продукции на (рыночную) цену p_0 этой продукции.

Издержками C фирмы называют общие выплаты фирмы в определённом временном периоде за все виды затрат $C = p_1 x_1 + p_2 x_2$, где x_1 и x_2 - объемы затрачиваемых (используемых) фирмой ресурсов (факторов производства), p_1 и p_2 - рыночные цены на эти ресурсы (факторы производства).

Прибылью PR фирмы в определённом временном периоде называется разность между полученным фирмой доходом R и ее издержками производства:

$$PR = R - C,$$

или

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

Последнее равенство есть выражение прибыли фирмы в терминах затрачиваемых (используемых) ресурсов. Напомним, что $y = f(x_1, x_2)$ - производственная функция фирмы, которая выражает общий объем y выпускаемой фирмой продукции через объемы x_1 и x_2 затрачиваемых (используемых) ресурсов.

В теории фирмы принято считать, что если фирма функционирует в условиях чистой (совершенной) конкуренции, на рыночные цены p_0, p_1 и p_2 она влиять не может. Фирма *"соглашается"* с ценами p_0, p_1 и p_2. Случаи функционирования фирмы в условиях чистой монополии, монополистической конкуренции и олигополии специально рассматриваются в рамках курса по микроэкономике.

Основная *цель* фирмы заключается в *максимизации* прибыли путем рационального *распределения* затрачиваемых (используемых) *ресурсов*. Формально задача максимизации прибыли в определённом временном периоде имеет вид: $PR \to \max$. Такая постановка задачи максимизации зависит от того, какой конкретно временной промежуток (долговременный или краткосрочный) предшествует периоду, в котором фирма максимизирует свою прибыль.

В случае долговременного промежутка фирма может свободно выбирать любой вектор $x = (x_1, x_2)$ затрат из пространства затрат (формально из неотрицательного ортанта $x_1 \geq 0$, $x_2 \geq 0$ плоскости Ox_1x_2), поэтому задача максимизации прибыли в случае *долговременного* промежутка имеет следующий вид:

$$p_0f(x_1, x_2) - (p_1x_1 + p_2x_2) = PR(x_1, x_2) \to \max$$

при условии, что

$$x_1 \geq 0, x_2 \geq 0$$

(постановка задачи в терминах затрачиваемых (используемых) ресурсов).

В случае краткосрочного промежутка фирма должна учитывать неизбежные лимиты на объемы затрачиваемых (используемых) ею ресурсов, которые формально могут быть записаны в виде нелинейного, вообще говоря, неравенства

$$g(x_1, x_2) \leq b$$

(ограничений вида $g(x_1, x_2) \leq b$ может быть несколько). Следовательно, задача максимизации прибыли для *краткосрочного* промежутка имеет вид задачи математического программирования:

$$p_0f(x_1, x_2) - (p_1x_1 + p_2x_2) = PR(x_1, x_2) \to \max$$

при условии, что

$$g(x_1, x_2) \leq b,$$
$$x_1 \geq 0, x_2 \geq 0$$

(постановка задачи в терминах затрачиваемых (используемых) ресурсов).

Линия уровня функции $z = p_1x_1 + p_2x_2$ издержек производства называется *изокостой* (см. рис. 11.1).

В связи с тем, что по экономическому смыслу $x_1 \geq 0$, $x_2 \geq 0$ (ибо x_1 и x_2 - это объемы затрачиваемых (используемых) ресурсов), строго говоря, изокоста есть отрезок прямой, попадающий в неотрицательный ортант плоскости Ox_1x_2. Таким образом изокосты - это отрезки A_0B_0, A_1B_1, A_2B_2,... (см. рис. 11.1). Отрезки A_0B_0, A_1B_1, A_2B_2 параллельны. Отрезок A_1B_1, расположенный "северо-восточнее" отрезка A_0B_0, соответствует большим издержкам производства. Сле-

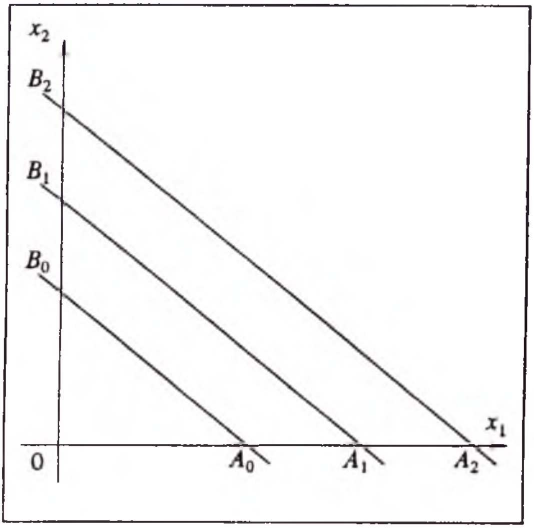


Рис. 11.1

довательно, если для отрезка A_2B_2 издержки производства C равны величине C_2, т.е. $C = C_2$, для отрезка A_1B_1 издержки производства $C = C_1$, для отрезка A_0B_0 издержки производства $C = C_0$, то $C_0 < C_1 < C_2$. Верно и обратное, т.е. если $C_0 < C_1 < C_2$, то отрезок A_2B_2, соответствующий издержкам производства C_2, расположен "северо-восточнее" параллельного ему отрезка A_1B_1, соответствующего издержкам производства C_1. Аналогично, отрезок A_1B_1 расположен "северо-восточнее" параллельного ему отрезка A_0B_0, соответствующего издержкам производства C_0. Для отрезка A_0B_0 имеем следующее аналитическое представление:

$$C_0 = p_1x_1 + p_2x_2,\ x_1 \geq 0,\ x_2 \geq 0,$$

для отрезка A_1B_1:

$$C_1 = p_1x_1 + p_2x_2,\ x_1 \geq 0,\ x_2 \geq 0,$$

для отрезка A_2B_2:

$$C_2 = p_1x_1 + p_2x_2,\ x_1 \geq 0,\ x_2 \geq 0.$$

11.2. Функции спроса на факторы (ресурсы) в случае долговременного промежутка

В связи с тем, что, как правило, $f(x_1,0)=f(0,x_2)=0$ (т.е. если хотя бы один ресурс не затрачивается (не используется), то объем выпускаемой продукции равен нулю), экономически осмысленными являются векторы (x_1,x_2) затрат ресурсов, для которых $x_1>0$, $x_2>0$. Поэтому в случае *долговременного* промежутка задача максимизации прибыли представляет собой обычную задачу на *глобальный* абсолютный максимум при $x_1>0$ и $x_2>0$. Из математического анализа известно, что точки *локального* абсолютного максимума следует искать только среди точек (x_1,x_2), которые удовлетворяют системе уравнений

$$\frac{\partial PR(x_1,x_2)}{\partial x_1}=0,\quad \frac{\partial PR(x_1,x_2)}{\partial x_2}=0,$$

или, в развернутом виде (ибо прибыль $PR(x_1, x_2) = p_0 f(x_1, x_2) - (p_1x_1 + p_2x_2)$),

$$p_0\frac{\partial f(x_1,x_2)}{\partial x_1}=p_1,\quad p_0\frac{\partial f(x_1,x_2)}{\partial x_2}=p_2. \tag{1}$$

Если вторые частные производные производственной функции $f(x_1,x_2)$ при всех $x_1>0$, $x_2>0$ таковы, что

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2}<0,\quad \frac{\partial^2 (x_1,x_2)}{\partial x_1^2}\cdot\frac{\partial^2 f(x_1,x_2)}{\partial x_2^2}-\left(\frac{\partial^2 f(x_1,x_2)}{\partial x_1\partial x_2}\right)^2>0$$

(эти неравенства несколько сильнее условий 2'' и 3 раздела 11.2 главы 10, посвященной производственным функциям), то график производственной функции $y=f(x_1,x_2)$ в трехмерном пространстве Ox_1x_2y есть поверхность, выпуклая вверх. Следовательно, график прибыли $PR(x_1, x_2)$, получаемый путем вычитания из графика функции $p_0f(x_1,x_2)$ плоскости $y = p_1x_1 + p_2x_2$, являющейся графиком издержек производства, имеет вид "шапочки", у которой есть "макушка". "Макушка" соответствует *глобальному* максимуму прибыли:

$$PR(x_1, x_2) = p_0f(x_1, x_2) - (p_1x_1 + p_2x_2).$$

Из этого геометрического факта следует, что система (1) имеет единственное *решение* (x_1^0, x_2^0), которое является *точкой* не только

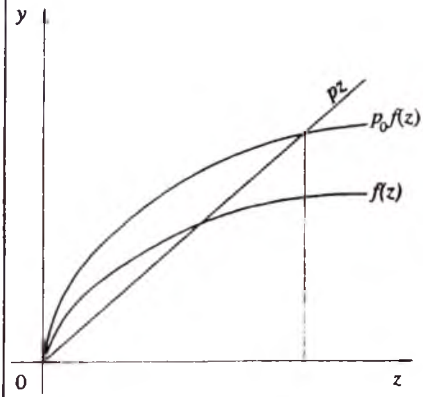


Рис. 11.2а

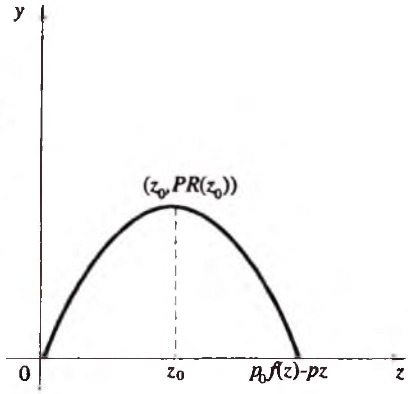


Рис. 11.2б

локального, но и *глобального* (искомого нами) максимума прибыли $PR(x_1, x_2)$. Вектор (x_1^0, x_2^0) затрат ресурсов, который является *решением задачи* максимизации прибыли $PR(x_1, x_2) = p_0f(x_1, x_2) - (p_1x_1 + p_2x_2)$, называется *локальным* (частичным) *рыночным равновесием* фирмы (в случае долговременного промежутка).

Рисунок графика прибыли $PR(x_1, x_2)$ в трехмерном пространстве, вообще говоря, достаточно сложен. Поэтому график прибыли представим схематически на плоскости Ozy, где координатная ось Oz изображает плоскость Ox_1x_2. На рис. 11.2а даны графики производственной функции $f(z)$, дохода фирмы $p_0f(z)$ и издержек производства pz. На рис. 11.2б изображен график прибыли $PR(z) = p_0f(z) - pz$, который получен вычитанием из графика дохода фирмы $p_0f(z)$ графика издержек производства pz. Точка $(z_0, PR(z_0))$ есть "макушка" "шапочки" - графика функции $PR(z) = p_0f(z) - pz$.

Подставив вектор (x_1^0, x_2^0) в уравнение (1), получим тождества

$$p_0\frac{\partial f(x_1^0, x_2^0)}{\partial x_1} = p_1,\ p_0\frac{\partial f(x_1^0, x_2^0)}{\partial x_2} = p_2, \tag{2}$$

откуда путем почленного деления первого тождества на второе получаем

$$\frac{\dfrac{\partial f(x_1^0, x_2^0)}{\partial x_1}}{\dfrac{\partial f(x_1^0, x_2^0)}{\partial x_2}} = \frac{p_1}{p_2}, \tag{3}$$

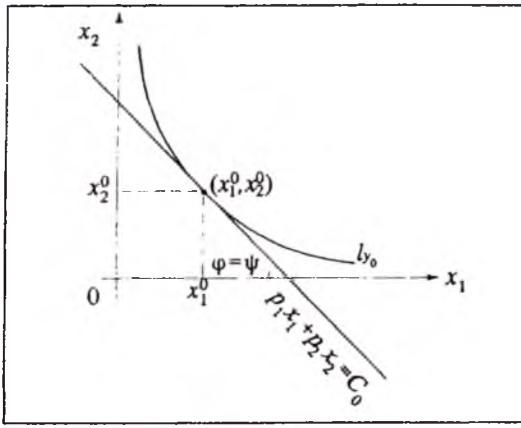


Рис. 11.3

т.е. в точке (x_1^0, x_2^0) локального рыночного равновесия фирмы отношение предельной производительности первого ресурса к предельной производительности второго ресурса равно отношению рыночных цен на эти ресурсы.

Проведем через точку (x_1^0, x_2^0) изокванту и изокосту, которые эту точку содержат. Уравнение изокванты имеет вид $f(x_1, x_2) = y_0$, где $y_0 = f(x_1^0, x_2^0)$. Уравнение изокосты имеет вид $p_1x_1 + p_2x_2 = C_0$, где $C_0 = p_1x_1^0 + p_2x_2^0$. Перепишем уравнение $f(x_1, x_2) = y_0$, выразив явно переменную x_2 через переменную x_1, т.е. в виде $x_2 = h(x_1)$ (обратим внимание, что уравнения $f(x_1,x_2) = y_0$ и $x_2 = h(x_1)$ формально разные, но они аналитически описывают одну и ту же изокванту - см. рис. 11.3).

$$\operatorname{tg}\varphi = -\frac{dh(x_1^0)}{dx_1} = \left[\frac{\partial f(x_1^0,x_2^0)}{\partial x_1}\right] \Big/ \left[\frac{\partial f(x_1^0,x_2^0)}{\partial x_2}\right]. \qquad (4)$$

Из математического анализа известно, что для изокосты $p_1x_1 + p_2x_2 = C_0$ отношение $\frac{p_1}{p_2} = \operatorname{tg}\psi$. Из (3), (4) и (5) следует, что $\operatorname{tg}\varphi = \operatorname{tg}\psi$, что означает, что касательная K к изокванте в точке (x_1^0, x_2^0) совпадает с изокостой, т.е. в точке (x_1^0, x_2^0) изокванта обязательно касается изокосты K (см. рис. 11.3). Получена важная (подчеркиваем это особо!) геометрическая характеристика локального рыночного равновесия (x_1^0, x_2^0) фирмы - *касание* в этом равновесии изокванты и изокосты.

Отметим, что, приступая к решению задачи максимизации прибыли, мы не имели конкретных изокванты и изокосты, которые касаются друг друга в точке (x_1^0, x_2^0), ибо не имели самой этой точки. Касающиеся друг друга изокванта и изокоста появляются после того, как аналитически найдено локальное рыночное равновесие (x_1^0, x_2^0) путем решения системы уравнений (1).

Левая ("четырехэтажная") дробь в (3) есть не что иное, как $R_{12}(x_1^0,x_2^0)$ - предельная норма замены первого ресурса вторым в точке (x_1^0,x_2^0).

Равенство (3) выражает следующий фундаментальный факт теории фирмы:

в точке локального рыночного равновесия (x_1^0,x_2^0) предельная норма замены $R_{12}(x_1^0,x_2^0)$ первого ресурса вторым равна отношению $\frac{p_1}{p_2}$ рыночных цен на эти ресурсы.

Поскольку x_1^0 и x_2^0 получаются в виде решения системы уравнений (1), постольку x_1^0 и x_2^0 есть функции цен (p_0, p_1, p_2), т.е.

$$x_1^0=d_1(p_0,p_1,p_2),\ x_2^0=d_2(p_0,p_1,p_2). \tag{5}$$

Выражения (5) называются *функциями спроса на ресурсы (затраты)*. Их значения x_1^0 и x_2^0 выражают оптимальные выборы затрат (использования) ресурсов как функции цены выпускаемой продукции и цен на ресурсы.

Подставив функции (5) в производственную функцию $y = f(x_1, x_2)$, получим выражение

$$y^0=f(d_1(p_0,p_1,p_2),\ d_2(p_0,p_1,p_2))=s(p_0,p_1,p_2),$$

которое называется *функцией предложения выпуска*.

Функции спроса на ресурсы и функция предложения выпуска являются однородными нулевой степени по всем своим аргументам p_0, p_1 и p_2, т.е. $d_1(tp_0, tp_1, tp_2) = d_1(p_0, p_1, p_2)$, $d_2(tp_0, tp_1, tp_2) = d_2(p_0, p_1, p_2)$, $s(tp_0, tp_1, tp_2) = s(p_0, p_1, p_2)$ для любого числа $t > 0$. Свойство однородности означает, что одновременное изменение всех цен p_0, p_1, p_2 в одно и то же число раз t (т. е. при изменении масштаба, но не структуры цен) не меняет x_1^0, x_2^0 и y^0, что важно с содержательной точки зрения. С математической точки зрения однородность нулевой степени функции спроса и функции предложения является простым фактом, ибо максимизация прибыли $PR(x_1,x_2) = tp_0f(x_1,x_2) - (tp_1x_1 + tp_2x_2)$ сводится к системе уравнений (2), поскольку на множитель $t > 0$ можно сократить.

11.3. Функции спроса на факторы (ресурсы) в случае краткосрочного промежутка

В случае *краткосрочного* промежутка рассмотрим конкретный пример, когда второй ресурс фирма может использовать только в объеме, равном $x_2^\# > 0$. Тогда задача максимизации прибыли превращается в задачу максимизации функции одной переменной:

$$PR(x_1,x_2^\#)=p_0f(x_1,x_2^\#)-(p_1x_1+p_2x_2^\#),$$

и вместо системы уравнений (1) появляется только одно уравнение

$$\frac{\partial PR(x_1,x_2^*)}{\partial x_1} = 0, \text{ или } p_0\frac{\partial f(x_1,x_2^*)}{\partial x_1} = p_1. \qquad (6)$$

Как и в разделе 11.2, считаем (исходя из содержательных экономических соображений), что уравнение (6) имеет единственное решение $x_1 = x_1^\#$ ($x_1^\#$ зависит от $x_2^\#$, т.е. $x_1^\# = x_1(x_2^\#)$), следовательно, в случае краткосрочного промежутка локальное рыночное равновесие есть вектор $(x_1^\#,x_2^\#)$. С помощью рис. 11.4 дадим ему наглядную геометрическую интерпретацию. Если бы объем $x_2^\#$ второго ресурса не был лимитирован, то, как видно из рис. 11.4, локальным рыночным равновесием была бы точка касания (x_1^0,x_2^0), в которой тот же объем выпускаемой продукции (для точек $(x_1^\#,x_2^\#)$ и (x_1^0,x_2^0) изокванта одна и та же) получился бы при меньших издержках производства (изокоста, содержащая точку $(x_1^\#,x_2^\#)$, расположена "северо-восточнее" изокосты, содержащей точку (x_1^0,x_2^0)). В точке $(x_1^\#,x_2^\#)$ локального рыночного равновесия содержащие ее изокванта и изокоста пересекаются, но *не* касаются. В рассматриваемом случае на самом деле $x_1^\#= x_1^\#(x_2^\#,p_0,p_1,p_2)$, это и есть *функция спро-*

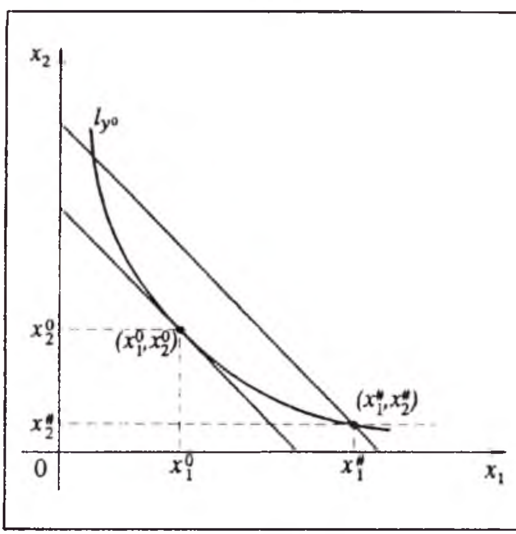


Рис. 11.4

са на первый ресурс при фиксированном объёме x_2^* второго ресурса. *Функция предложения выпуска* имеет вид $y=f(x_1^*(x_2^*,p_0,p_1,p_2),x_2^*)$. Может случиться так, что точки (x_1^*,x_2^*) и (x_1^0,x_2^0) сольются в одну, и тогда получится та ситуация, которая уже была проанализирована в разделе 11.2.

11.4. Комбинация ресурсов (факторов производства), максимизирующая объём выпуска при ограничении на затраты

Для случая *долговременного* промежутка рассмотрим задачу максимизации объема выпускаемой продукции при ограничении затрат на приобретение ресурсов (факторов), т.е. рассмотрим задачу 1:

$$f(x_1, x_2) \to \max \tag{7}$$

при условии, что

$$p_1x_1 + p_2x_2 \le V, \tag{8}$$
$$x_1 \ge 0,\ x_2 \ge 0. \tag{9}$$

Величина V не обязательно равна величине C_0 (см. раздел 11.2). Решение этой задачи математического программирования допускает наглядную геометрическую интерпретацию (см. рис. 11.5). Ограни-

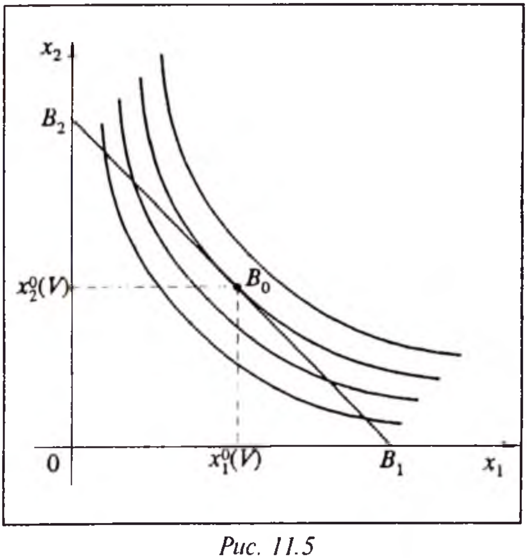


Рис. 11.5

чениям (8) и (9) соответствует треугольник OB_1B_2 плоскости Ox_1x_2. Максимизация функции (7) геометрически соответствует тому, что мы переходим на все более "северо-восточные" изокванты, пока они имеют еще общие точки с треугольником OB_1B_2 (прямая B_1B_2 имеет уравнение $p_1x_1 + p_2x_2 = V$). Изокванты - гладкие линии, выпуклые к точке O (а это так, ибо $f(x_1, x_2)$ не произвольная функция двух переменных, а ***производственная функция,*** т.е. функция, удовлетворяющая определенным требованиям гладкости и выпуклости), поэтому решению задачи (7), (8), (9) соответствует изокванта, которая касается гипотенузы (изокосты) B_1B_2 в точке B_0. Любая изокванта, расположенная "северо-восточнее" этой изокванты, содержащей точку B_0, не подходит, ибо не имеет общих точек с треугольником OB_1B_2. Координаты $x_1^0(V)$ и $x_2^0(V)$ точки B_0 и дают решение задачи (8), (9), (10).

В связи с тем, что это решение $(x_1^0(V), x_2^0(V))$ обращает ограничение (8) в равенство $p_1x_1^0 + p_2x_2^0 = V$, вместо задачи (7), (8), (9) можно рассмотреть более простую задачу на условный экстремум

$$f(x_1, x_2) \to \max \tag{7}$$

при наличии ограничения

$$\begin{gathered} p_1x_1 + p_2x_2 = V \\ (x_1 \geq 0,\ x_2 \geq 0), \end{gathered} \tag{10}$$

заданного в виде равенства.

Задачи (7), (8), (9) и (7), (10) разные, но решение $(x_1^0(V), x_2^0(V))$ у них одно и то же (в случае отсутствия так называемых угловых решений, которые здесь не рассматриваются). Поскольку сумма $p_1x_1+p_2x_2$ равна издержкам производства, постольку целесообразно заменить V на C и формально перейти к задаче максимизации объема выпускаемой продукции для случая *долговременного* промежутка при *фиксированных издержках производства C* (величина C играет роль параметра и не обязательно равна величине C_0 (см. раздел 2)):

$$f(x_1, x_2) \to \max \tag{7}$$

при условии, что

$$\begin{gathered} p_1x_1 + p_2x_2 = C \\ (x_1 \geq 0,\ x_2 \geq 0). \end{gathered} \tag{11}$$

Геометрическое решение задачи (7), (11) также наглядно очевидно (см. рис. 11.6): следует переходить на все более "северо-восточ-

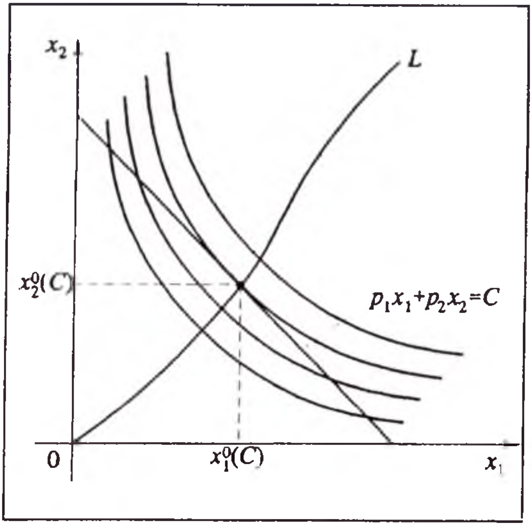


Рис. 11.6

ные" изокванты до тех пор, пока они продолжают иметь общие точки с изокостой, соответствующей фиксированным издержкам производства C. Ясно, что решением задачи максимизации выпуска будет точка $(x_1^0(C), x_2^0(C))$ касания последней из допустимых изоквант и фиксированной изокосты $p_1x_1 + p_2x_2 = C$. Эта точка касания зависит от величины издержек производства C (поэтому и написано $(x_1^0(C), x_2^0(C))$. Если издержки производства C изменятся, то изменится, вообще говоря, и точка $(x_1^0(C), x_2^0(C))$. Множество точек $(x_1^0(C), x_2^0(C))$, соответствующих различным значениям C, образуют линию L (см. рис. 11.6), которая называется *долговременной линией развития фирмы*. Точка (x_1^0,x_2^0) локального рыночного равновесия фирмы (см. раздел 11.2) обязательно принадлежит линии L.

Решим задачу (7), (11) формально с помощью функции Лагранжа

$$L(x_1,x_2,\lambda) = f(x_1,x_2) + \lambda(C - p_1x_1 - p_2x_2).$$

Для функции Лагранжа выписываем систему уравнений

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial x_1}=0,\quad \frac{\partial L(x_1,x_2,\lambda)}{\partial x_2}=0,\quad \frac{\partial L(x_1,x_2,\lambda)}{\partial \lambda}=0$$

или в развернутом виде

$$\frac{\partial f(x_1,x_2)}{\partial x_1}=\lambda p_1,\quad \frac{\partial f(x_1,x_2)}{\partial x_2}=\lambda p_2,\quad C-p_1x_1-p_2x_2=0. \qquad (12)$$

Критическая точка $(x_1^0(C),x_2^0(C),\lambda^0(C))$ функции Лагранжа, удовлетворяющая системе (12) и взятая без последней координаты (множителя Лагранжа) $\lambda^0(C)$, т.е. точка $(x_1^0(C),x_2^0(C))$, и есть *решение задачи* (7), (11) максимизации выпуска при данных фиксированных издержках производства C. Подставив точку $(x_1^0(C),(x_2^0(C),\lambda^0(C))$ в первые два уравнения системы (12), получим два тождества. Поделив почленно первое тождество на второе, получим, очевидно, выражение (3) (множитель Лагранжа $\lambda^0(C)$ сократится). Получили аналитическое обоснование того, что в точке $(x_1^0(C),x_2^0(C))$ изокванта и изокоста *касаются* (см. рис. 11.6). Вообще говоря, критическая точка функции Лагранжа, взятая без последней координаты, не обязана быть решением задачи (7), (11) на условный максимум. В случае же производственной функции $f(x_1,x_2)$, удовлетворяющей определенным требованиям гладкости и выпуклости, критическая точка функции Лагранжа (без последней координаты) есть решение задачи (7), (11) на условный экстремум. Отметим также, что в случае производственной функции $f(x_1,x_2)$ $x_1^0(C)>0$, $x_2^0(C)>0$, $\lambda^0(C)>0$.

В разделе 11.2 в точке локального рыночного равновесия (x_1^0,x_2^0) были определены издержки производства $C_0 = p_1x_1^0 + p_2x_2^0$. Если в ограничении (12) положить $C = C_0$, то несложно показать, что $x_1^0(C_0) = x_1^0$, $x_2^0(C_0) = x_2^0$, а также $\dfrac{1}{\lambda^0(C_0)} = p_0$, т.е. величина, обратная множителю Лагранжа $\lambda^0(C_0)$, равна рыночной цене p_0 единицы выпускаемой фирмой продукции. Таким образом, предложена естественная экономическая интерпретация множителя Лагранжа $\lambda^0(C_0)$.

Подставив $x_1^0(C)$, $x_2^0(C)$ в выражение $y = f(x_1, x_2)$, получим, что

$$y = f(x_1^0(C),\ x_2^0(C)) = F(C), \tag{13}$$

т. е. получим $y = F(C)$ как функцию издержек производства C, а не как функцию $y = f(x_1,x_2)$ объемов x_1 и x_2 затрачиваемых (используемых) ресурсов. Выражение (13) называется *значением задачи* (7), (11).

Так построенная функция $y = F(C)$ соответствует случаю *долговременного* промежутка.

Имея функцию $y = F(C)$, можно выписать выражение для прибыли в терминах *издержек производства* $PR(C) = p_0F(C) - C$ (сравните с выражением для прибыли фирмы в терминах *затрачиваемых (используемых) ресурсов* - см. раздел 11.1).

Задача максимизации объема выпускаемой продукции при *фиксированных издержках производства* C для случая *краткосрочного* промежутка, когда лимитирован объем $x_2^\#$ второго ресурса, имеет вид

$$f(x_1,\ x_2^\#) \to \max \tag{14}$$

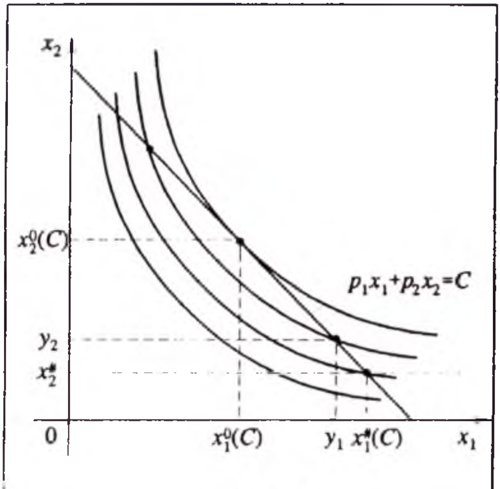


Рис. 11.7

при условии, что

$$p_1x_1 + p_2x_2^{\#} = C \qquad (15)$$
$$(x_1 \geq 0).$$

Ограничимся наглядным геометрическим решением задачи (7), (14) (см. рис. 11.7).

Перемещаемся по изоквантам на "северо-восток" до того момента, пока изокванта не пройдет через точку $(x_1^{\#}(C), x_2^{\#})$, которая и есть решение задачи (14), (15). От этой изокванты далее на "северо-восток" идти нельзя, ибо в точках $y = (y_1, y_2)$ пересечения новых изоквант и фиксированной изокосты $p_1x_1 + p_2x_2 = C$ имеет место неравенство $y_2 \neq x_2^{\#}$ (см. рис. 11.7).

Если бы условия $x_2 = x_2^{\#}$ не было, то решением задачи максимизации объема выпускаемой продукции была бы точка $(x_1^0(C), x_2^0(C))$ (см. рис. 7), которая соответствует случаю долговременного промежутка. Очевидно, $f(x_1^0(C), x_2^0(C)) > f(x_1^{\#}(C), x_2^{\#})$, ибо изокванта, проходящая через точку $(x_1^0(C), x_2^0(C))$, расположена "северо-восточнее" изокванты, проходящей через точку $(x_1^{\#}(C), x_2^{\#})$.

Получен важный *результат теории фирмы:*

при одних и тех же издержках производства C объем выпускаемой продукции для *долговременного* промежутка больше (точнее не меньше) объема выпускаемой продукции для *краткосрочного* промежутка.

Эти объемы сравняются, если издержки производства C^* будут такими, что $x_2^0(C^*) = x_2^{\#}$ (см. рис. 11.8). Горизонтальная прямая $x_2 = x_2^{\#}$ называется *краткосрочной линией развития производства фирмы.*

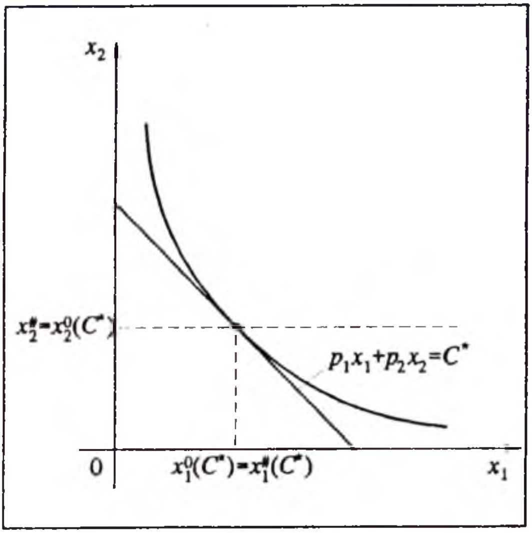


Рис. 11.8

11.5. Комбинация ресурсов (факторов производства), минимизирующая издержки при фиксированном (общем) объёме выпуска

Для случая *долговременного* промежутка рассмотрим задачу минимизации издержек производства при *фиксированном объеме y* выпускаемой продукции (т.е. рассмотрим задачу 2):

$$p_1x_1+p_2x_2=C(x_1,x_2) \to \min \tag{16}$$

при условии, что

$$y=f(x_1,x_2) \quad (x_1 \geq 0,\ x_2 \geq 0). \tag{17}$$

Геометрически решение задачи (16), (17) аналогично решению задачи (7), (10) (см. рис. 11.9): в случае задачи (16), (17) следует перемещаться по изокостам на "юго-запад" (ибо имеем задачу минимизации) до тех пор, пока они продолжают иметь общие точки с изоквантой, соответствующей фиксированному объему y. Ясно, что решением задачи минимизации издержек будет общая точка $(x_1^0(y),x_2^0(y))$ изокосты и фиксированной изокванты. Эта точка *касания* зависит от объема y (поэтому и написано $(x_1^0(y),x_2^0(y)))$. Если объем y изменится, то изменится и точка $(x_1^0(y),x_2^0(y))$. Множество

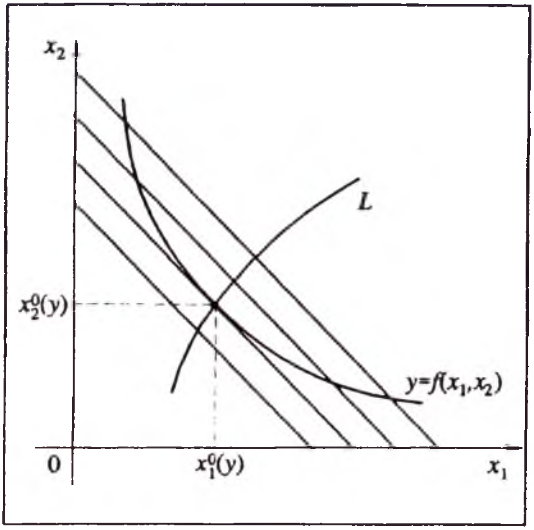


Рис. 11.9

точек $(x_1^0(y),x_2^0(y))$, соответствующих различным объемам y выпускаемой продукции, образуют линию L (см. рис. 11.9), которая, очевидно, совпадает с линией L рис. 11.6.

Решим задачу (16), (17) формально с помощью функции Лагранжа

$$L(x_1,x_2,\lambda) = p_1x_1 + p_2x_2 + \lambda(y - f(x_1,x_2)).$$

Для функции Лагранжа выписываем систему уравнений

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial x_1}=0,\quad \frac{\partial L(x_1,x_2,\lambda)}{\partial x_2}=0,\quad \frac{\partial L(x_1,x_2,\lambda)}{\partial \lambda}=0$$

или в развернутом виде

$$p_1=\lambda\frac{\partial f(x_1,x_2,\lambda)}{\partial x_1},\quad p_2=\lambda\frac{\partial f(x_1,x_2,\lambda)}{\partial x_2}=0 \qquad (18)$$

Критическая точка $(x_1^0(y),\ x_2^0(y),\ \lambda^0(y))$ функции Лагранжа, удовлетворяющая системе (18) и взятая без последней координаты $\lambda^0(y)$, т.е. точка $(x_1^0(y),\ x_2^0(y))$, и есть *решение задачи* (16), (17) минимизации издержек при данном фиксированном объеме производства y. Подставив точку $(x_1^0(y),x_2^0(y),\lambda^0(y))$ в первые два уравнения системы

(18), получим два тождества. Поделив почленно первое тождество на второе, получим, очевидно, выражение (3) (множитель $\lambda^0(y)$ сократится, как и в случае раздела 4 множитель $\lambda^0(y)>0$). Получили аналитическое обоснование того, что изокоста и изокванта *касаются* в точке $(x_1^0(y),x_2^0(y))$ (см. рис. 9). Характер взаимосвязи между критической точкой функции Лагранжа без последней координаты и решением задачи (16), (17) минимизации может быть прокомментирован здесь подобно тому, как это было сделано в разделе 4 для задачи максимизации.

В разделе 2 в точке локального рыночного равновесия (x_1^0,x_2^0) был определён объём производства $y_0=f(x_1^0,x_2^0)$. Если в ограничении (17) положить $y=y_0$, то несложно показать, что $x_1^0(y_0)=x_1^0$, $x_2^0(y_0)=x_2^0$, а также $\lambda^0(y_0)=p_0$, т.е. множитель Лагранжа $\lambda^0(y_0)$ равен рыночной цене p_0 единицы выпускаемой продукции. Таким образом, предложена естественная экономическая интерпретация множителя Лагранжа $\lambda^0(y_0)$.

Подставив $x_1^0(y)$ и $x_2^0(y)$ в выражение $C(x_1, x_2) = p_1x_1 + p_2x_2$, получим выражение для издержек

$$p_1x_1^0(y) + p_2x_2^0(y) = C(y)$$

как функцию объема выпускаемой продукции, а не как функцию $C(x_1, x_2) = p_1x_1 + p_2x_2$ объемов затрачиваемых (используемых) ресурсов. Выражение $C(y) = p_1x_1^0(y) + p_2x_2^0(y)$ называется *значением задачи* (16), (17). Так построенная функция издержек $C(y)$ соответствует случаю *долговременного промежутка*. Имея выражение $C(y)$, выпишем в явном виде представление прибыли в виде функции *объемов y выпускаемой продукции:*

$$PR(y) = p_0y - C(y)$$

(сравните с выражением для прибыли фирмы в терминах *затрачиваемых (используемых) ресурсов* - см. раздел 11.1). Выражение $PR(y) = p_0y - C(y)$ играет важную роль в микроэкономике.

Пусть $(x_1^0(C), x_2^0(C))$ и $y = h(C)$ есть *решение и значение* задачи максимизации (7), (11) (см. раздел 11.4).

Пусть $(x_1^0(y), x_2^0(y))$ и $C = C(y)$ есть *решение и значение* задачи минимизации (16), (17) (см. начало этого раздела 11.5).

Пусть значение C в (12) равно значению $C(y)$ задачи минимизации (16), (17). Тогда значение задачи максимизации (7), (11) (см. раздел 11.4) будет равно y.

Наоборот, пусть значение y в (17) равно значению $y = h(C)$ задачи максимизации (7), (11). Тогда значение задачи минимизации (16), (17) равно C.

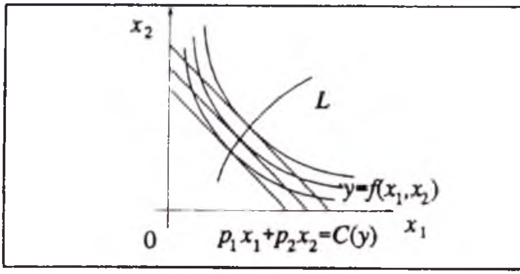


Рис. 11.10

Таким образом, наблюдается взаимная зависимость задач (7), (11) и (16), (17) (см. рис. 11.10).

Задача минимизации издержек производства при *фиксированном объеме y* выпускаемой продукции для случая *краткосрочного* промежутка, когда фиксирован объем $x_2^{\#}$ второго ресурса, имеет вид (y играет роль параметра)

$$p_1x_1 + p_2x_2^{\#} = C(x_1, x_2^{\#}) \text{ (min)} \tag{19}$$

при условии, что

$$y = f(x_1, x_2) \\ (x_1 \geq 0). \tag{20}$$

Ограничимся наглядным геометрическим решением задачи (19), (20) (см. рис. 11.11).

Имеет место важный результат теории фирмы:

при одном том же объеме y выпускаемой продукции издержки производства для случая *долговременного* промежутка меньше (точнее не больше) издержек производства для случая *краткосрочного* промежутка. Эти издержки производства равны друг другу, если объем y^* производства будет таким, что $x_1^0(y^*) = x_2^{\#}$ (см. рис. 11.12).

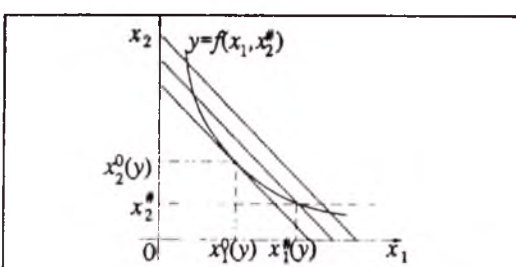


Рис. 11.11

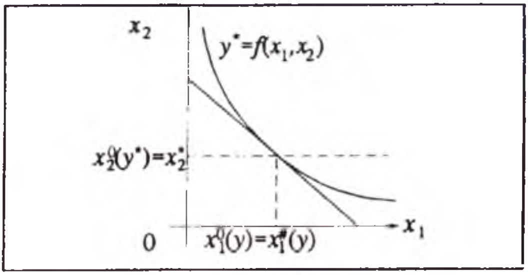


Рис. 11.12

Вопросы к главе 11

1. Сформулируйте определения исходных понятий теории оптимизации производства.
2. Сформулируйте основную цель функционирования фирмы.
3. В чем состоит принципиальное отличие (в содержательных и в формальных терминах) в постановке задачи максимизации прибыли фирмы для случаев долговременного и краткосрочного временных промежутков?
4. Сформулируйте определение изокосты. Дайте экономическую интерпретацию следующему геометрическому факту: фиксированная точка (v_1, v_2) $(v_1 > 0,\ v_2 > 0)$ принадлежит прямой, имеющей уравнение $p_1x_1 + p_1x_2 = C_0$.
5. Пусть два набора (v_1, v_2) и (w_1, w_2) затрачиваемых (используемых) ресурсов таковы, что для набора (v_1, v_2) издержки производства меньше, чем для набора (w_1, w_2). Как расположены относительно друг друга изокосты, соответствующие наборам (v_1, v_2), (w_1, w_2)? Является ли обязательным одновременное выполнение двух неравенств $w_1 > v_1$, $w_2 > v_2$?
6. Опишите взаимное расположение изокосты и изокванты, проходящих через точку (x_1^0, x_2^0) локального рыночного равновесия фирмы (случай долговременного промежутка). Дайте обоснование их взаимному расположению.
7. Опишите взаимное расположение изокосты и изокванты, проходящих через точку локального рыночного равновесия фирмы (случай краткосрочного промежутка).
8. Чему равна (предельная технологическая) норма замены одного ресурса (для определенности, первого) другим (для определенности, вторым) в точке локального рыночного равновесия в случае долговременного промежутка?
9. Что такое функция спроса на ресурсы (затраты)?
10. Что такое функция предложения выпуска?

11. Сформулируйте задачу максимизации объема выпускаемой продукции при ограничении затрат на приобретение ресурсов (факторов) на аналитическом языке и на геометрическом языке (в случаях долговременного и краткосрочного промежутков).
12. Опишите аналитическое (для случая долговременного промежутка) и геометрическое (для случаев долговременного и краткосрочного промежутков) решение задачи вопроса 11.
13. Сформулируйте задачу минимизации издержек производства при фиксированном объеме выпускаемой продукции на аналитическом языке и на геометрическом языке (для случаев долговременного и краткосрочного промежутков).
14. Опишите аналитическое (для случая долговременного промежутка) и геометрическое (для случаев долговременного и краткосрочного промежутков) решение задачи вопроса 13.
15. Опишите взаимосвязь между задачами вопросов 11 и 13 (в случае долговременного промежутка).
16. Как определяется долговременная линия расширения производства фирмы?
17. Как определяется краткосрочная линия расширения производства фирмы?

ГЛАВА 12

ЭКОНОМИЧЕСКАЯ ДИНАМИКА И ЕЕ МОДЕЛИРОВАНИЕ

Задачи, решаемые экономической наукой и практикой, делятся, в зависимости от учета фактора времени, на статические и динамические. Статика изучает состояния экономических объектов, относящиеся к определенному моменту или периоду времени, без учета изменения их параметров во времени. В динамических задачах отражается не только зависимость переменных от времени, но и их взаимосвязи во времени. Например, динамика инвестиций определяет динамику величин основного капитала, что в свою очередь является важнейшим фактором изменения объема выпуска.

Время в экономической динамике может рассматриваться как непрерывное или дискретное. Непрерывное время удобно для моделирования, так как позволяет использовать аппарат дифференциального исчисления и дифференциальных уравнений. Дискретное время удобно для приложений, поскольку статистические данные всегда дискретны и относятся к конкретным единицам времени. Для дискретного времени может использоваться аппарат разностных уравнений. Заметим, что большинство известных моделей экономической динамики существуют как в непрерывном, так и в дискретном вариантах. В обоих вариантах для них могут быть получены, как правило, аналогичные результаты, и уровень сложности самих моделей примерно одинаков.

12.1. Показатели экономической динамики

Показатели, характеризующие динамику экономического объекта, - это абсолютные приросты, темпы роста и прироста.

Если рассматривается зависящая от времени величина $A(t)$, то абсолютный прирост от момента 0 до момента 1 равен $\Delta A(1) = A(1) - A(0)$, дискретный темп роста $\eta_1 = \frac{A(1)}{A(0)}$, дискретный темп прироста $\alpha_1 = \eta_1 - 1 = \frac{A(1) - A(0)}{A(0)}$. Отметим, что в англоязычной литературе термином “growth rate” (“темп роста”) называют обычно показатель $\alpha = \eta - 1$, то есть темп прироста в нашей терминологии.

Если темп прироста α неизменен во времени, то динамика показателя $A(t)$ может быть описана как $A(t) = A(0)\,(1+\alpha)^t$.

Если величина $A(t)$ есть непрерывная функция времени, то рост ее с постоянным темпом записывается как $A(t) = A(0)\,e^{\lambda t}$, где $e \approx 2{,}72$

- основание натуральных логарифмов, а λ - непрерывный темп прироста, который в общем случае рассчитывается как $\lambda(t) = \frac{d(A(t))}{A(t)\cdot dt}$ или $\lambda(t) = \frac{A_t'}{A(t)} = \frac{\dot{A}}{A}$. Величина $dA(t)=A_t'\cdot dt = \dot{A}(t)\cdot dt$ - дифференциал (главная линейная часть приращения) $A(t)$, где $A_t'=\dot{A}(t)$ - производная функции $A(t)$ по времени. При росте величины $A(t)$ с непрерывным темпом прироста λ дискретный темп роста $\frac{A(t+1)}{A(t)}$ равен e^{λ}, что при малых λ близко к $(1+\lambda)$, то есть к темпу роста при дискретном темпе прироста λ.

Рассмотрим величины темпов прироста для сумм и произведений показателей.

Пусть показатель $S(t)$ есть сумма $A(t)$ и $B(t)$, растущих соответственно, с постоянными непрерывными темпами α и β, причем $\alpha>\beta$. Тогда

$$S(t) = A(t) + B(t) = A(0)\cdot e^{\alpha t} + B(0)\cdot e^{\beta t} = \\ = A(0)\cdot e^{\alpha t}\cdot\left[1 + e^{(\beta-\alpha)t}\cdot\frac{B(0)}{A(0)}\right] \qquad (1)$$

Поскольку $(\beta-\alpha)<0$, величина в квадратных скобках стремится к единице, и темп прироста суммы приближается к темпу быстрее растущего составляющего, то есть к α.

Пусть величина $P(t)$ есть произведение $A(t)$ и $B(t)$ с непрерывными темпами прироста α и β. В этом случае:

$$P(t) = A(t)\cdot B(t) = A(0)\cdot e^{\alpha t}\cdot B(0)\cdot e^{\beta t} = P(0)\cdot e^{(\alpha+\beta)t}, \qquad (2)$$

то есть темп прироста произведения равен сумме темпов прироста сомножителей. Если α и β - дискретные темпы прироста $A(t)$ и $B(t)$, то

$$P(t)=A(t)\cdot B(t)=A(0)\cdot(1+\alpha)^t\cdot B(0)\cdot(1+\beta)^t=P(0)\cdot(1+\alpha+\beta+\alpha\beta)^t, \qquad (3)$$

При малых α и β величина $\alpha\beta$ пренебрежимо мала, и темп прироста произведения приближенно равен сумме темпов прироста сомножителей. Если же произведение $\alpha\,\beta$ значительно, то темп прироста произведения не может приближенно считаться равным сумме темпов прироста сомножителей, поскольку существенно ее превышает.

Связь объемных и темповых величин легко продемонстрировать на примере производственной функции (это уже сделано в главе 10 для частного случая ПФ Кобба-Дугласа). Пусть $Y(t)$, $K(t)$, $L(t)$ - объемные показатели выпуска, капитала и труда (непрерывные функции времени), а $y(t)$, $k(t)$, $l(t)$ - непрерывные темпы их прироста. Объемная ПФ с нейтральным техническим прогрессом (при постоянном темпе последнего, равном γ) имеет вид

$$Y(t) = f[K(t),\ L(t)]\ e^{\gamma t} \tag{4}$$

Логарифмируя эту зависимость, получаем:

$$\ln\ Y(t) = \ln f[K(t),\ L(t)] + \gamma t. \tag{5}$$

Далее дифференцируем по времени:

$$d\ln\ Y(t) = \frac{dY(t)}{Y(t)} = \frac{dY(t)}{dK(t)} \cdot \frac{K(t)}{Y(t)} \cdot \frac{dK(t)}{K(t)} + \frac{dY(t)}{dL(t)} \cdot \frac{L(t)}{Y(t)} \cdot \frac{dL(t)}{L(t)} + \gamma \mathrm{d}t,$$

то есть,

$$y(t) = \alpha(t)\ k(t) + \beta(t)\ l(t) + \gamma, \tag{6}$$

где $\alpha(t)$ и $\beta(t)$ - эластичности выпуска по капиталу и труду соответственно.

Эта линейная формула характеризует вклад темпов прироста факторов производства в общие темпы прироста дохода, а показатель γ характеризует вклад технического прогресса.

12.2. Понятие динамического равновесия в экономике. Простейшая модель равновесия

В экономической теории важным является понятие равновесия, то есть такого состояния объекта, которое он сохраняет при отсутствии внешних воздействий. Задачи экономической динамики включают как описание процессов выхода к состоянию равновесия, так и процессов трансформации самого этого состояния под воздействием внешних сил. Рассмотрим простую экономическую систему в состоянии равновесия и опишем движение такой системы в непрерывном и дискретном случаях. В первом случае динамика системы описывается с помощью дифференциального уравнения, во втором - разностного уравнения.

Дифференциальное уравнение связывает изменения показателя (пусть наша система описывается одним показателем $x(t)$, или просто x) со скоростью его движения x'_t, или $\dot{x}$. Будем считать, что скорость изменения показателя x пропорциональна величине его отклонения от равновесного значения x_e. Иными словами, чем дальше показатель отклонился от равновесного значения, тем быстрее он стремится вернуться к нему. Если в уравнении присутствует только первая производная x по времени, а сама связь линейна, то это линейное дифференциальное уравнение. Пусть оно имеет, например, следующий вид:

$$\dot{x} = k(x - x_e),$$

где k-коэффициент. В этом уравнении kx_e - свободный член; без него уравнение $\dot{x}=kx$ называется однородным и его общее решение $x = c\,e^{kt}$. Исходное неоднородное уравнение имеет частное решение $x = x_e$ (если величина x находится в состоянии равновесия), а общее его решение есть сумма любого частного решения и общего решения однородного уравнения, то есть $x = x_e + c\,e^{kt}$. Учитывая, что при $t = 0$ величина x равна $x(0)$, получаем $c = x(0)-x_e$, и $x(t) = x_e+(x(0)-x_e)e^{kt}$. Проверьте в качестве упражнения, что это решение удовлетворяет исходному уравнению. Если $k < 0$, то $e^{kt} \to 0$ и равновесие устойчиво, то есть при отклонении величины $x(t)$ от значения x_e она вновь стремится принять это значение. При $k > 0$ величина $e^{kt} \to \infty$ и, соответственно, $x(t)$ стремятся к бесконечности (если начальное состояние не совпадает с состоянием равновесия).

Система выходит к состоянию x_e, как это показано на рисунке 12.1а. Ее поведение при $k > 0$ показано на рис. 12.1б. Поведение динамических систем может также описываться, например, графиками рис. 12.1в - 12.1г. Поведение в дискретном времени может быть описано с помощью разностного уравнения, связывающего величины x в соседние моменты времени, то есть x_t и x_{t-1}. Например,

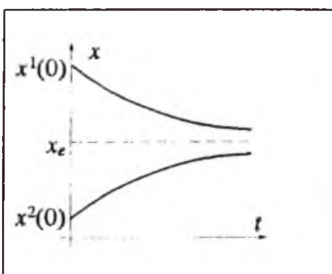


Рис. 12.1а

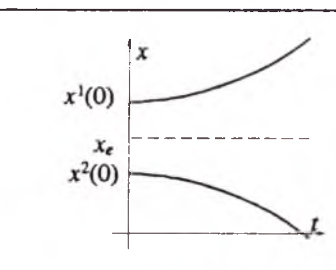


Рис. 12.1б

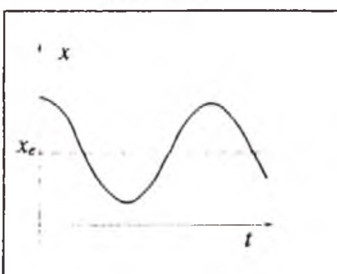


Рис. 12.1в

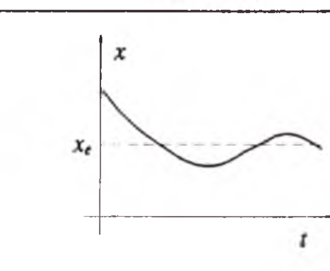


Рис. 12.1г

в дискретной ситуации, аналогичной уже описанной, может использоваться разностное уравнение $x_t = x_{t-1} + k(x_{t-1} - x_e)$, решением которого (проверьте!) является $x_t = x_e + (x(0) - x_e)\,(1+k)^t$. Это решение может быть найдено (аналогично непрерывному случаю) как сумма общего решения $x_t = c\,(1+k)^t$ для однородного уравнения $x_t = (1+k)\,x_{t-1}$, и частного решения $x_t = x_e$ для исходного разностного уравнения; с учетом $x_t = x(0)$ при $t = 0$. При $k < 0$ система в случае отклонения от x_e будет двигаться в направлении x_e, при $k > 0$ - уходить еще дальше от него. Равновесие устойчиво при $-2<k<0$ и неустойчиво при $k>0$ или k -2 (при $k<-1$ показатель x каждый раз "перескакивает" равновесное значение x_e, причем при $k<-2$ - слишком далеко, чтобы приблизиться в конце концов к x_e).

12.3. Примеры моделей экономической динамики

Рассмотрим теперь два примера моделей макроэкономической динамики, реализующих дискретный и непрерывный подходы. В обоих случаях модели носят весьма общий, абстрактный характер. В то же время их решение может быть найдено в явном виде, причем из него вытекают важные особенности для различных частных случаев соотношения их параметров. На этих моделях удобно продемонстрировать простейший аппарат дискретного и непрерывного динамического моделирования, проиллюстрировать важнейшие категории и проблемы макроэкономической динамики.

Паутинообразная модель

Эта модель позволяет исследовать устойчивость цен и объемов товаров на рынке, описываемом традиционными кривыми спроса и предложения (они показаны на рис. 12.2) при наличии запаздывания во времени (лага).

Пусть производители (например, зерновая ферма) определяют предложение товара в текущем периоде на основе цен, установив-

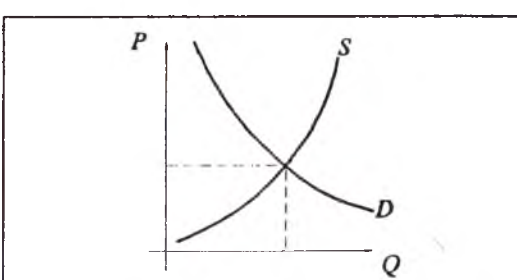


Рис. 12.2

шихся в предшествующем периоде, то есть $Q^s(t) = S_t(p_{t-1})$. Таким образом, в функцию предложения вклинивается временной лаг продолжительностью в одну единицу времени. Действительно, решение об объеме производства принимается с учетом текущих цен, но производственный цикл имеет определенную продолжительность, и соответствующее этому решению предложение появится на рынке по окончании данного цикла.

Кривая спроса характеризует зависимость объема спроса на товар от цены товара в данном периоде, то есть $Q^D(t) = D_t(p_t)$. Таким образом, динамику цены можно описать системой уравнений

$$\{Q_t^s = S_t(p_{t-1}),\ Q_t^D = D_t(p_t),\ Q_t^D = Q_t^s\}$$

или одним уравнением

$$D_t(p_t) = S_t(p_{t-1}). \tag{7}$$

Из этого уравнения можно найти значение цены p_t в текущий момент времени по известному значению p_{t-1} в предшествующий момент времени. Схема решения очень проста:

$$Q_0 \to p_0 = D^{-1}(Q_0) \to Q_1 = S(p_0) \to p_1 = D^{-1}(Q_1) \to Q_2 = S(p_1) \to \ldots$$

(где D^{-1} - обратная функция спроса).

В качестве частного случая рассмотрим паутинообразную модель, в которой функции спроса и предложения линейны:

$$S(p) = A + Bp_{t-1},\ D(p) = C - Ep_t,\ S(p) = D(p). \tag{8}$$

Здесь $B>0$, так как функция предложения возрастающая; $E>0$, так как функция спроса убывающая; $C>A>0$, то есть $D(0)>C(0)>0$ (считаем, что при нулевой цене спрос превышает предложение). Уравнение, описывающее динамику такой системы, имеет вид

$$D(p_t) = S(p_{t-1}), \text{ или } C - E\,p_t = A + B\,p_{t-1}.$$

Найдем сначала равновесную цену p^* и равновесный объем производства Q^*. Они должны удовлетворять уравнениям

$$Q^* = C - Ep^* = A + Bp^*,$$

откуда

$$p^* = \frac{C - A}{B + E} \text{ и } Q^* = \frac{BC + AE}{B + E}.$$

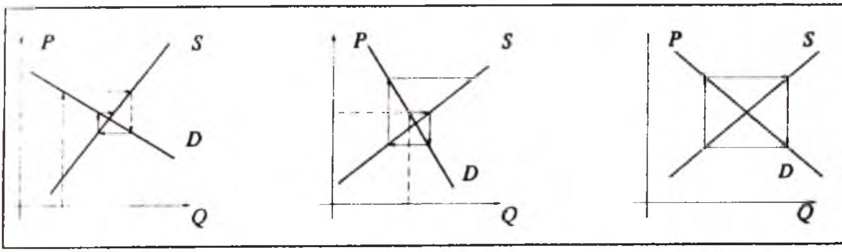


Рис. 12.3а Рис. 12.3б Рис. 12.3в

Далее необходимо исследовать поведение цен и объемов производства в том случае, если начальная точка не совпадает с равновесной. Вначале эту задачу можно решить графически, получив рисунок типа "паутины", подтверждающий ее название. Задав некоторое первоначальное количество товара и цену, не совпадающие с точкой равновесия, будем последовательно наносить точки в соответствии с процедурой расчета по модели, соединяя их горизонтальными или вертикальными прямыми линиями. Из графического анализа можно получить следующие результаты. Если кривая предложения наклонена круче, чем кривая спроса, то равновесие на таком рынке будет устойчивым (см. рис. 12.3а). Если кривая спроса наклонена круче, чем кривая предложения, то равновесие на рынке будет неустойчивым (см. рис. 12.3б). Наконец, при равном наклоне кривых спроса и предложения цены на рынке будут испытывать регулярные колебания с постоянной амплитудой (см. рис. 12.3в).

Теперь перейдем к формальному анализу модели. Выражая p_t через p_{t-1}, имеем следующее рекуррентное соотношение $p_t = \frac{C - A}{E} - \frac{B}{E}$. Последовательно применяя это соотношение, находим

$$p_1 = \frac{C - A}{E} - \frac{B}{E} \cdot p_0;\ p_2 = \frac{C - A}{E} - \frac{B}{E} \cdot \left(\frac{C - A}{E} - \frac{B}{E} \right) \cdot p_0$$

Или в общем виде

$$p_t = \frac{C - A}{E} \cdot \left[1 - \frac{B}{E} + \left(\frac{B}{E} \right)^2 + \ldots + (-1)^{t-1} \left(\frac{B}{E} \right)^{t-1} \right] + (-1)^t \left(\frac{B}{E} \right)^t \cdot p_0$$

Выражение в скобках есть сумма геометрической прогрессии:

$$S_n = a_1 \cdot (1 + q + q^2 + \ldots + q^{n-1}) = a_1 \frac{1 - q^n}{1 - q}.$$

Если $|q|<1$, то $\lim\limits_{n \to \infty} S_n = \frac{a_1}{1 - q}$.

Для паутинообразной модели $q = -\frac{B}{E}$, $a_1 = \frac{C - A}{E}$. Отсюда получаем выражение для цены p_t в произвольный момент времени t:

$$p_t = \frac{C - A}{E} \cdot \frac{1 - (-1)^t \left[\frac{B}{E}\right]^t}{1 + \frac{B}{E}} + (-1)^t \left[\frac{B}{E}\right]^t \cdot p_0 . \qquad (10)$$

Очевидно при $\frac{B}{E} < 1$ $\left[\frac{B}{E}\right]^t \to 0$ и $p_t \to \frac{C - A}{B + E} = p^*$, то есть при более крутом наклоне кривой предложения, чем кривой спроса, равновесие является устойчивым. Если $\frac{B}{E} > 1$, то есть более крутой является кривая спроса, то $\left[\frac{B}{E}\right]^t \to \infty$ и процесс расходится (равновесие неустойчиво). При $\frac{B}{E} = 1$, то есть при $B = E$, значения p_t чередуются вокруг равновесного значения.

Итак, определяющим моментом для устойчивости системы является менее сильная, сглаживающая реакция на изменения цены той функции, которая имеет временной лаг (здесь - функция предложения).

В реальности при $\frac{B}{E} > 1$ бесконечно возрастающих колебаний, конечно, не будет, так как при больших отклонениях от равновесия линейное приближение становится нереалистичным. В более реалистической нелинейной модели устанавливаются нелинейные колебания большой, но конечной амплитуды, которые являются прообразом экономических циклов подъема и спада производства.

Самостоятельно предлагается рассмотреть следующую задачу: предположим, что временной лаг, равный 1, присутствует не в функции предложения, а функции спроса: $S_t = A + Bp_t$; $D_t = C - Ep_{t-1}$; $S_t = D_t$. Каким станет условие сходимости к равновесной точке? Изобразить этот процесс графически.

12.4. Модели макроэкономической динамики

Модель Харрода-Домара

В качестве примера модели с непрерывным временем рассмотрим модель макроэкономической динамики (простейший ее вариант - модель Харрода-Домара). Модель описывает динамику дохода $Y(t)$, который рассматривается как сумма потребления $C(t)$ и инвестиций $I(t)$. Экономика считается закрытой, поэтому чистый экспорт равен нулю, а государственные расходы в модели не выделяются. Основная предпосылка модели роста - формула взаимосвязи между инвестициями и скоростью роста дохода. Предполагается, что скорость роста дохода пропорциональна инвестициям: $I(t) = B\frac{dY}{dt}$, где B - коэффициент капиталоемкости прироста дохода, или приростной капиталоемкости (соответственно, обратная ему величина $\frac{1}{B}$ называется приростной капиталоотдачей). Тем самым в модель фактически включаются следующие предпосылки:

- инвестиционный лаг равен нулю: инвестиции мгновенно переходят в прирост капитала. Формально это означает, что $\Delta K(t) = I(t)$, где $\Delta K(t)$ - непрерывная функция прироста капитала во времени;
- выбытие капитала отсутствует;
- производственная функция в модели линейна; это вытекает из пропорциональности прироста дохода приросту капитала:

$$dY(t) = \frac{1}{B}d(K(t))dt.$$

Линейная производственная функция $Y(t) = a\,L(t) + b\,K(t) + c$, где $b = \frac{1}{B}$, обладает этим свойством в том случае, если либо a=0, либо $L(t)$=const. Тем самым следующая предпосылка такова:

- затраты труда постоянны во времени либо выпуск не зависит от затрат труда, поскольку труд не является дефицитным ресурсом;
- модель не учитывает технического прогресса.

Перечисленные предпосылки, конечно, существенно огрубляют описание динамики реальных макроэкономических процессов, делают затруднительным применение данной модели, например, для непосредственного расчета или прогноза величины совокупного выпуска или дохода. Однако данная модель и не предназначена для этого; в то же время ее относительная простота позволяет более глубоко изучить взаимосвязь динамики инвестиций и роста выпуска, получить точные формулы траекторий рассматриваемых параметров при сделанных предпосылках.

Зависимость, связывающая между собой во времени показатели инвестиций, определяемый ими объем основного капитала и уровень выпуска (дохода), является базовой во всех моделях макроэкономической динамики. Кроме того, в этих моделях необходимо определить принципы формирования структуры выпуска (дохода), распределения его между составляющими, прежде всего - между потреблением и накоплением. Эти принципы могут основываться на оптимизационном подходе (обычно это максимизация совокупных объемов потребления в той или иной форме), экстраполяционном, равновесном и других. В рассматриваемой модели предполагается, что динамика объема потребления $C(t)$ задается экзогенно. Этот показатель может считаться постоянным во времени, расти с заданным постоянным темпом или иметь какую-либо другую динамику (в первых двух случаях более просто получить решение модели).

Простейший вариант модели получается, если считать $C(t) = 0$. Этот случай совершенно нереалистичен с практической точки зрения, однако в нем все ресурсы направляются на инвестиции, в результате чего могут быть определены максимальные технически возможные темпы роста. В этом случае получаем:

$$Y(t) = C(t) + I(t) = 0 + B\frac{dY(t)}{dt} = B\dot{Y}(t). \qquad (11)$$

Это - линейное однородное дифференциальное уравнение, и его решение имеет вид $Y(t) = Y(0)\cdot e^{\left(\frac{1}{B}\right)t}$ (что легко проверить дифференцированием). Непрерывный темп прироста здесь равен $\frac{1}{B}$. Это максимально возможный (технологический) темп прироста.

Пусть теперь $C(t) = C$ постоянно во времени. Получаем неоднородное линейное дифференциальное уравнение $Y(t)=B\dot{Y}(t)+C$. Его частным решением является $Y(t) = C$, и складывая его с общим решением однородного уравнения $Y(t) = A\cdot e^{\left(\frac{1}{B}\right)t}$, получаем его общее решение $Y(t) = A\cdot e^{\left(\frac{1}{B}\right)t} + C$, откуда, подставив $t=0$, имеем $A = Y(0)-C = I(0)$ и $Y(t) = (Y(0) - C)\cdot e^{\left(\frac{1}{B}\right)t} + C$. Непрерывный темп прироста дохода $y(t) = \frac{\dot{Y}(t)}{Y(t)}$ в этом решении равен $y(t) = \frac{1}{B}\cdot\left[1 - \frac{C}{Y(t)}\right]$. Он составляет $\frac{1}{B}\cdot\left[1 - \frac{C(0)}{Y(0)}\right]$ в начальный момент времени (при $t = 0$) и, возрастая, стремится к $\frac{1}{B}$ при $t \to \infty$ (что понятно, поскольку доход растет, а постоянный объем потребления составляет все меньшую его долю). Величина в скобках $\alpha(t) = \left[1 - \frac{C}{Y(t)}\right]$ есть норма накопления в момент времени t, и темп прироста дохода оказывается пропорциональным этой величине, как и показателю приростной капиталоотдачи $\frac{1}{B}$.

Итак, при прочих равных рост нормы накопления пропорционально увеличивает темпы прироста дохода. В то же время это снижает уровень текущего потребления, и для разрешения проблемы согласования конкурентных целей увеличения темпов роста и уровня текущего благосостояния в модель обычно включают элементы оптимизации. В этом случае решается оптимизационная задача на максимум общего объема потребления за конечный или бесконечный период времени. Для отражения предпочтительности более раннего получения результата в модель включается временное дисконтирование, при котором более ранний результат учитывается в критерии с большим "весом".

Наконец, рассмотрим вариант модели с показателем потребления $C(t)$, растущим с постоянным темпом r: $C(t) = C(0)\ e^{rt}$. Дифференциальное уравнение этой модели имеет вид $Y(t)=B\dot{Y}(t)+C(0)\cdot e^{rt}$. Решение этого уравнения (проверьте дифференцированием!) таково:

$$Y(t) = \left[Y(0) - \frac{C(0)}{1 - Br}\right]\cdot e^{\frac{t}{B}} + \left[\frac{C(0)}{1 - Br}\right]\cdot e^{rt}. \tag{12}$$

Из общих соображений ясно, что темп прироста потребления r не должен быть больше максимально возможного общего темпа прироста $\frac{1}{B}$, так как иначе потребление будет занимать все большую и в конце концов - подавляющую часть дохода, что сведет к нулю сначала инвестиции, а затем и доход. Ясно это и из формулы решения модели, поскольку в случае $r > \frac{1}{B}$ коэффициент $\frac{1}{1 - Br}$ отрицателен, а e^{rt} растет быстрее, чем $e^{\left(\frac{1}{B}\right)t}$, - следовательно, второе слагаемое при этом отрицательно и через некоторое время "перевесит" первое.

В решении рассматриваемой модели роста при $r < \frac{1}{B}$ многое зависит от соотношения между r и $\rho_0 = \frac{\alpha_0}{B}$ (в числителе стоит $\alpha_0 = 1 - \frac{C(0)}{Y(0)}$ - норма накопления в начальный момент времени t=0). Если $r = \rho_0$, то темп прироста дохода равен темпу прироста потребления, и решением является $Y(t) = Y(0)\cdot e^{\left(\frac{1}{B}\right)t}$. Норма накопления $\alpha(t)$ в этом случае постоянна во времени и равна α_0, а темп прироста дохода пропорционален норме накопления и обратно пропорционален приростной капиталоемкости. Именно эта модификация модели экономического роста, в которой постоянна норма накопления, называется моделью Харрода-Домара.[1]

Если в рассматриваемой модели роста $\frac{1}{B} > r > \rho_0$, то требуемый темп прироста потребления оказывается слишком высоким для экономики. В этом случае коэффициент $\left[Y(0) - \frac{C(0)}{1 - Br}\right]$ отрицателен и, поскольку $\frac{1}{B} > r$, первое, отрицательное слагаемое в решении "перевешивает" в конце концов второе. Поэтому темп прироста дохода падает и становится с некоторого момента отрицательным. Через некоторое время сам доход становится равным нулю, после чего модель теряет экономический смысл. Это аналогично случаю $r \geq \frac{1}{B}$, хотя здесь уже дело не в том, что нужный темп прироста потребления в принципе недостижим за длительный период. В данном случае слишком низкой оказывается начальная норма накопления α_0.

Если $r < \rho_0$, то норма накопления, а вместе с ней и темп прироста дохода растут, причем последний в пределе приближается к $\frac{1}{B}$. Однако в этом случае происходит "накопление ради накопления", ибо потребление растет заданным темпом r, а темп прироста дохода удается увеличить за счет более быстрого роста инвестиций. Норма накопления α_0 здесь превышает Br, и если исходить из задачи мак-

---

[1] Кстати, в хорошо известной нашим экономистам модели роста - схемах расширенного воспроизводства К. Маркса - получается аналогичный результат. Темп прироста дохода в схемах К.Маркса равен $\frac{n_1 z}{1 + h_1}$, где n_1 - норма накопления в первом подразделении, $z = \frac{M}{V}$ - норма прибавочной стоимости, $h_1 = \frac{C_1}{V_1}$ - органическое строение капитала в первом подразделении (C - постоянный капитал, V - переменный капитал, M - прибавочная стоимость).

В терминах рассматриваемой модели роста:

$$B = \frac{\Delta(C + V)}{\Delta(V + M)} = \frac{\Delta V(1 + h_1)}{\Delta V(1 + z)} = \frac{1 + h_1}{1 + z}, \text{ а}$$

$$\alpha_0 = \frac{\Delta V + \Delta C}{V + M} = \frac{\Delta V + \Delta C}{M\left(1 + \frac{1}{z}\right)} = n_1 \frac{z}{z + 1}.$$

Отсюда $\frac{\alpha_0}{B} = \frac{n_1 z(1 + z)}{(1 + z)(1 + h_1)} = \frac{n_1 z}{1 + h_1}$, что показывает совпадение результатов в этих двух моделях роста с постоянными значениями структурных параметров.

симизации объема потребления, то эта норма слишком высока. Более высокий ее уровень требует увеличения инвестиций $I(0)$ за счет сокращения потребления $C(0)$ в начальный момент, что при фиксированном *темпе прироста* потребления r обусловливает более низкий его *уровень* на всей траектории. В то же время нужный темп прироста потребления $r < \frac{1}{B}$ можно поддерживать, как показано выше, при $\alpha_0 = Br$. Таким образом, если требуется поддерживать постоянный темп прироста потребления r, не превышающий технологического темпа, то для максимизации объема потребления за любой период нужно установить начальную норму накопления $\alpha_0 = Br$. Более сложен вопрос о том, какой уровень темпа r более предпочтителен. Большая его величина позволяет обеспечить больший объем потребления за длительный период, но это происходит за счет сокращения потребления на начальном этапе. Таким образом, для выбора значения r (если оно предполагается постоянным) нужна информация о межвременных предпочтениях лица, принимающего решение.

Модель Солоу

Другой тип модели экономического роста представляет модель, предложенная лауреатом Нобелевской премии Р.Солоу. По сравнению с уже рассмотренной моделью роста модель Солоу позволяет более точно описать некоторые особенности макроэкономических процессов. Во-первых, производственная функция в этой модели нелинейна и обладает свойством убывания предельной производительности. Во-вторых, модель учитывает выбытие основного капитала. В-третьих, в модель Солоу включается описание динамики трудовых ресурсов и технического прогресса и их влияние на экономический рост. В-четвертых, здесь ставится и решается задача максимизации уровня потребления на некотором множестве устойчивых траекторий. Все это, конечно, усложняет структуру модели, и получение точных формул для траекторий изменения основных ее показателей становится существенно более сложной задачей. Поэтому некоторые другие аспекты описываются в базовой модели Солоу упрощенно: например, считаются постоянными норма сбережений и норма выбытия капитала, инвестиционные лаги отсутствуют, а производственная функция имеет постоянную отдачу от масштаба. Кроме того, на начальном уровне анализа модели ищутся не траектории изменения всех ее показателей (как в модели Харрода-Домара), а характеристики состояний устойчивого равновесия, к которым система выходит в долгосрочном периоде. С формальной точки зрения это представляет собой существенно более простую задачу.

Мы не ставим здесь задачу подробно излагать модель Солоу, сформулируем лишь основные ее предпосылки, обозначения и выводы.

Предпосылки и обозначения модели Солоу:

- производственная функция имеет вид $Y = F(K,L)$ (Y - выпуск или доход, K - капитал, L - труд). Отдача от масштаба постоянна: $F(zK, zL) = zF(K,L)$. Предельная производительность факторов положительна, но убывает:

$$Y'_K>0;\ Y'_L>0;\ Y''_{KK}<0;\ Y''_{LL}<0;$$

- величина выбытия капитала W пропорциональна его величине K:

$$W = \delta \cdot K,$$

где δ - норма выбытия;
- норма сбережений (инвестиций) α постоянна, и инвестиции I равны $\alpha \cdot Y$;
- доход Y распределяется на потребление и инвестиции: $Y = C + I$;
- численность занятых L растет с постоянным темпом n;
- трудосберегающий технический прогресс имеет темп g, то есть число единиц труда с постоянной эффективностью в расчете на одного работающего растет с темпом g.

При сделанных предпосылках производственную функцию можно рассматривать как зависимость производительности труда $y = \frac{Y}{L}$ от его капиталовооруженности $k = \frac{K}{L}$: $y=f(k)$ (здесь L - число единиц труда с постоянной эффективностью (то есть численность занятых при отсутствии трудосберегающего технического прогресса, либо численность условных работников с одинаковой эффективностью - при его наличии). Это вытекает из того, что $Y = F(K,L) = L \cdot F\left(\frac{K}{L},1\right)$ $= L \cdot F(k)$. Инвестиции приводят к росту капиталовооруженности, а выбытие капитала, рост численности работающих и числа единиц труда с постоянной эффективностью - к ее снижению. Прирост капиталовооруженности k в результате инвестиций равен $i=\frac{I}{L}$. Темп снижения капиталовооруженности за счет остальных факторов равен $(\delta+n+g)$ (в точности равен, если Y, K, L - непрерывные функции времени, и приближенно равен в дискретном случае при малых δ, n, g). Величина снижения капиталовооруженности за счет этих факторов равна $(\delta+n+g)\ k$.

Величина k находится в состоянии устойчивого равновесия, если ее прирост за счет инвестиций равен ее уменьшению за счет других факторов. Поскольку $Y=C+I$, после деления этого тождества на L имеем $y=c+i$, где y - доход, c - потребление, а i - инвестиции на одну единицу труда с постоянной эффективностью. Следовательно, величина i равна $\alpha f(k)$. Условие стабильности показателя k, таким образом, записывается как

$$(\delta+n+g) k^* = \alpha f(k^*), \tag{13}$$

и величина k^* называется устойчивым уровнем капиталовооруженности. На рис. 4 показана устойчивость равновесия при $k=k^*$. Это - точка равновесия для показателя k, поскольку в этой точке величина удельного прироста капиталовооруженности равна величине ее удельного сокращения, и показатель k остается неизменным. Это равновесие устойчиво, поскольку при $k_1<k^*$ удельные инвестиции превышают уменьшение капиталовооруженности, и ее величина растет. В случае $k_1<k^*$, наоборот, удельные инвестиции ниже, чем уменьшение капиталовооруженности, и ее величина падает, пока не достигнет k^*.

Из рис. 4 можно видеть, что в случае увеличения нормы сбережения α график функции инвестиций пойдет выше и, следовательно пересечет прямую $(\delta+n+g)$ k правее. Итак, рост нормы сбережения приводит к увеличению устойчивого уровня капиталовооруженности k^*, а следовательно, и устойчивого уровня дохода на единицу труда $y^*=f(k^*)$.

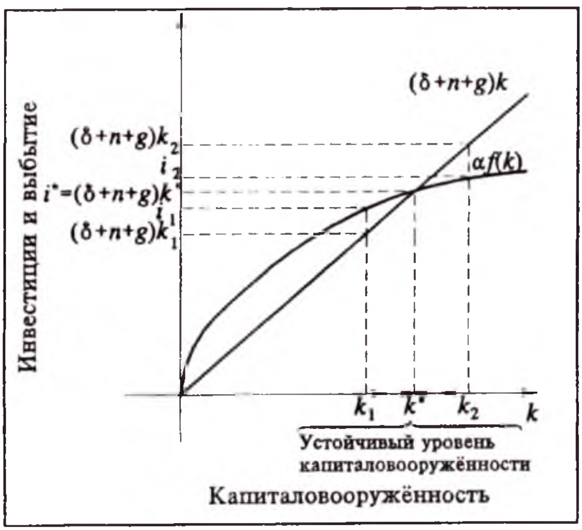


Рис. 4

Если численность работающих не растет (или растет медленнее), то есть показатель n равен нулю (или меньше по величине), то прямая $(\delta+n+g)\ k$ имеет меньший наклон и точка k^* сдвигается вправо. То же самое происходит при более низком (или нулевом) темпе трудосберегающего технического прогресса g.

В устойчивом состоянии темп прироста показателей k,y,c,i равен нулю. Поскольку все это - удельные показатели в расчете на единицу труда с постоянной эффективностью, а эффективность труда одного занятого растет с темпом g, показатели капитала, дохода, потребления и инвестиций в расчете на одного занятого растут с темпом g. При росте численности занятых с темпом n общий объем капитала, дохода, потребления и инвестиций растет в устойчивом состоянии с темпом $(n+g)$. Следовательно, модель Солоу показывает, что единственным источником длительного, устойчивого роста дохода на одного работника, а следовательно, и душевого потребления, является технический прогресс.

Как уже показано, каждому уровню нормы сбережения α соответствует определенное устойчивое состояние и свой уровень устойчивого потребления на единицу труда с постоянной эффективностью c^*. Можно поставить задачу отыскания устойчивого состояния, в котором величина c^* максимальна среди всех таких состояний. Поскольку в любом устойчивом состоянии выполняется равенство

$$i^* = (\delta+n+g)\ k^*,$$

и

$$c^* = y^* - i^* = f(k^*) - (\delta+n+g)\ k^*,$$

требуется максимизировать по k^* функцию $[f(k^*)-(\delta+n+g)\ k^*]$. Необходимым условием максимума дифференцируемой функции является равенство нулю ее производной; в данном случае это означает равенство

$$f'(k^*) = \delta + n + g. \tag{14}$$

Это правило выбора оптимального объема капитала для максимизации удельного объема потребления называется *Золотым правилом*. Соответствующая ему величина капиталовооруженности k^{**} называется капиталовооруженностью по Золотому правилу, а норма сбережения α^* - нормой сбережения по Золотому правилу. Она может быть найдена из уравнения $(\delta+n+g)\cdot k^{**} = \alpha^*\cdot f(k^{**})$, являющегося необходимым условием устойчивого состояния. Удельная величина потребления по Золотому правилу находится как разница между доходом и инвестициями:

$$c^{**} = f(k^{**}) - (\delta+n+g)\cdot k^{**}.$$

Рис. 12.5 иллюстрирует Золотое правило:

Рис. 12.5

В точке k^{**} касательная к графику производственной функции параллельна прямой $(\delta+n+g)\cdot k$. Если темп роста численности занятых более низок, либо более низок темп трудосберегающего технического прогресса, то прямая $(\delta+n+g)\cdot k$ становится более пологой, и точка k^{**} сдвигается вправо, а удельная величина потребления c^{**} растет. Статистика в целом подтверждает, что в странах с более быстрым ростом потребления уровень душевого потребления более низок, хотя, конечно, в каждой конкретной стране своя производственная функция, норма выбытия и исходное состояние развития (вовсе не обязательно устойчивое).

Если первоначальная величина капиталовооруженности k^* меньше, чем k^{**}, то имеет смысл увеличить норму сбережения до величины, соответствующей Золотому правилу, и постепенно экономика выйдет на максимальный уровень удельного потребления c^{**}. Отметим, однако, что вначале удельный уровень потребления снизится и лишь затем начнет постепенно расти, наряду с ростом удельных инвестиций и выпуска. Если же первоначальная величина капиталовооруженности k^* больше, чем k^{**}, то нужно снизить норму сбережения до уровня, соответствующего Золотому правилу. Тогда экономика также постепенно выйдет на уровень удельного потребления c^{**}. В этом случае вначале удельный уровень потребления вырастет, превысив c^{**}, и затем начнет постепенно снижаться к c^{**}, наряду со снижением удельных инвестиций (выросших в началь-

Рис. 12.6а

Рис. 12.6б

ный момент) и выпуска. Динамика показателей y, c, i для двух описанных случаев представлена на рис. 12.6а и 12.6б.

Таким образом, если нас интересует прежде всего рост потребления в ближайшей перспективе, а не максимальный его равновесный уровень в долгосрочном периоде, то и задача оптимизации должна быть сформулирована по-другому. Задачи максимизации потребления на ограниченном периоде времени также хорошо изучены, но мы не будем здесь на них останавливаться.

Вопросы и задачи к главе 12

1. В чем основное различие задач экономической статики и динамики? Как соотносится рассматриваемый период времени для статических и динамических задач?
2. В чем различие содержания решаемых задач, математического аппарата и получаемых результатов для экономических моделей с дискретным и непрерывным временем?
3. Как может быть записана динамика показателя, растущего:
   а) с постоянным дискретным темпом?
   б) с постоянным непрерывным темпом?
4. Стоимость ценной бумаги каждый следующий год повышается на 50% или понижается на 40%. Что произойдет с этой стоимостью в длительной перспективе?
5. Доход Y равен сумме потребления C и инвестиций I. Дискретный темп прироста потребления равен 10%, инвестиций - 25%. В начальном году ($t = 0$) C=500, I=150. Чему равен темп роста дохода Y в году 2?
6. Пусть в течение n лет, в конце каждого года, производится выплата кредитору суммы денег, равной 1. Процентная ставка равна i, все деньги кладутся кредитором в банк под этот процент. Какова будет у него сумма денег к концу года n? Какова должна быть ежегодная выплата в случае i=15%, n=5, чтобы накопленная сумма денег равнялась 100?
7. Как пересчитать непрерывные темпы прироста в дискретные и наоборот?
8. Пусть оценена производная функция Кобба-Дугласа в темповой записи $y = 0{,}4 \cdot k + 0{,}7 \cdot l + 0{,}5$, где y, k, l - годовые темпы прироста дохода, капитала и труда (в %):
   а) можно ли по этой формуле оценить величину вклада технического прогресса в темп прироста дохода?
   б) можно ли оценить долю вклада технического прогресса в темпы прироста дохода?
   в) как изменится формула, если темпы прироста измерять не в процентах, а в абсолютном выражении?
9. Сформулируйте понятие экономического равновесия. Чем устойчивое равновесие отличается от неустойчивого?
10. Для дифференциального уравнения $\dot{x}=k(x-x_e)$ приведите пример частного решения, сформулировав это понятие.
11. Запишите решение разностного уравнения $x_t = x_{t-1} + k(x_{t-1} - x_e)$ и поясните его поведение при разных значениях коэффициента k. Может ли решение этого уравнения, и если да, то в каком случае, "перескакивать" состояние равновесия x_e?
12. Имеется паутинообразная модель $S_t = 20 + 30p_{t-1}$; $D_t = 100 - 50p_t$; $S_t = D_t$. Пусть p_0=0,5, чему равно p_1?

13. Пусть в паутинообразной модели функция спроса равна $D_t = \frac{3}{p_t}$; функция предложения $S_t = 5p_{t-1}$, $p_0=1$. Изобразите графически динамику цен и объемов выпуска. Каковы равновесные цена и выпуск? Является ли равновесие устойчивым?
14. Пусть в макроэкономической модели роста темп прироста потребления задается экзогенно. Что произойдет, если он больше технологического темпа прироста $\frac{1}{B}$?
15. Как связан темп прироста выпуска с нормой накопления?
16. Как соотносятся темпы прироста выпуска в модели Харрода-Домара и схемах расширенного воспроизводства К.Маркса?
17. Как выбрать норму накопления при заданном темпе прироста потребления в макромодели роста?
18. В чем состоит проблема выбора наилучшего темпа роста потребления в модели Харрода-Домара? Какие Вы можете предложить подходы к ее решению?
19. Чем предпосылки модели Солоу отличаются от предпосылок модели Харрода-Домара? Какие общие принципы заложены в этих моделях?
20. Рассмотрите частный случай модели Солоу с постоянной численностью занятых и без технического прогресса. Запишите формулу для устойчивого состояния. Сформулируйте Золотое правило сбережения.
21. Рассмотрите частный случай модели Солоу с постоянной численностью занятых и с трудосберегающим техническим прогрессом с темпом g. Запишите формулу для устойчивого состояния. Сформулируйте Золотое правило сбережения.
22. Пусть производственная функция имеет вид $Y = 5 K^{1/3} L^{2/3} e^{0,03t}$. Норма выбытия капитала 0,08. Численность занятых растет на 2% в год. Норма сбережения 25%. Каков устойчивый уровень капиталовооруженности единицы труда с постоянной эффективностью? Каков устойчивый уровень удельного дохода, инвестиций, потребления? Соответствует ли данная норма сбережения Золотому правилу? Если нет, то какой она должна стать для этого? Каков устойчивый уровень удельного дохода, инвестиций, потребления по Золотому правилу?

ГЛАВА 13
ВВЕДЕНИЕ В МАТЕМАТИЧЕСКУЮ ТЕОРИЮ ИГР

13.1. Основные понятия теории игр

В лекциях по теории оптимизации рассматривались такие задачи принятия решений, когда выбор решения осуществлялся одним лицом. В подобных задачах рационального ведения хозяйства решение выбирается при предположении о том, что известны целевая функция,различные способы действия и ограничения. В данной главе рассматриваются задачи принятия решений в ситуациях с *несколькими* участниками, когда значение целевой функции для каждого из субъектов зависит и от решений, принимаемых всеми остальными участниками. Предметом теории игр являются такие ситуации, в которых важную роль играют конфликты и совместные действия.

Одна из характерных черт всякого общественного, социально - экономического явления состоит в множественности, многосторонности интересов и в наличии сторон, выражающих эти интересы. Классическими примерами здесь являются ситуации, где, с одной стороны, имеется один покупатель, с другой - продавец (ситуация монополия-монопсония), когда на рынок выходят несколько производителей, обладающих достаточной силой для воздействия на цену товара (ситуация олигополии, в том числе дуополии, если число таких участников равно двум). Более сложные ситуации подобного рода возникают, если имеются объединения или коалиции лиц, участвующих в столкновении интересов, например, в том случае, когда ставки заработной платы определяются союзами или объединениями рабочих и предпринимателей, при анализе результатов голосования в парламенте и т.п.

Конфликт может возникнуть также из различия целей, которые отражают не только несовпадающие интересы различных сторон, но и многосторонние интересы одного и того же лица. Например, разработчик экономической политики обычно преследует разнообразные цели, согласуя противоречивые требования, предъявляемые к ситуации (рост объемов производства, повышение доходов, снижение экологической нагрузки и т.п.). Конфликт может проявляться не только в результате сознательных действий различных участников, но и как результат действия тех или иных “стихийных сил” (случай так называемых “игр с природой”). Множество подобных примеров можно встретить в биологии, социологии, психологии, политологии, военном деле и т.д.

И наконец, примерами игр являются обычные игры: салонные, спортивные, карточные и др. Именно с анализа подобных игр начиналась математическая теория игр; они и по сей день служат прекрасным материалом для иллюстрации положений и выводов этой теории.

В итоге, всякая претендующая на адекватность математическая модель социально-экономического явления должна отражать присущие ему черты *конфликта*, т.е. описывать:

а) множество заинтересованных сторон (мы будем называть их *игроками*; в литературе по теории игр они именуются также субъектами, лицами, сторонами, участниками). В случае, если число игроков конечно, они различаются по своим номерам (1-й игрок и 2-й игрок в игре в орлянку или в случае дуополии) или по присваиваемым им именам (например, Продавец и Покупатель в ситуации монополия-монопсония);

б) возможные действия каждой из сторон, именуемые также *стратегиями* или *ходами*;

в) интересы сторон, представленные *функциями выигрыша (платежа)* для каждого из игроков.

В теории игр предполагается, что функции выигрыша и множество стратегий, доступных каждому из игроков, общеизвестны, т.е. каждый игрок знает свою функцию выигрыша и набор имеющихся в его распоряжении стратегий, а также функции выигрыша и стратегии всех остальных игроков, и в соответствии с этой информацией организует свое поведение.

Формализация содержательного описания конфликта представляет собой его математическую модель, которую называют *игрой*.

Теория игр впервые была систематически изложена Дж.фон Нейманом и О.Монгерштерном в 1944 г., хотя отдельные результаты были опубликованы еще в 20-х годах. Нейман и Моргенштерн написали оригинальную книгу, которая содержала главным образом экономические примеры, поскольку экономическому конфликту легче всего придать численную форму. Во время второй мировой войны и сразу после нее теорией игр серьезно заинтересовались военные, которые увидели в ней аппарат для исследования стратегических решений.Затем главное внимание снова стало уделяться экономическим проблемам. Сейчас ведется большая работа, направленная на расширение сферы применения теории игр.

13.2. Классификация игр

Различные виды игр можно классифицировать, основываясь на том или ином принципе: по числу игроков, по числу стратегий, по свойствам функций выигрыша, по возможности предварительных переговоров и взаимодействия между игроками в ходе игры.

В зависимости от *числа игроков* различают игры с *двумя, тремя* и более участниками. Весь материал, представленный в теории оптимизации, можно рассматривать как теорию игр с одним игроком. В принципе возможны также игры с *бесконечным* числом игроков.

Согласно другому принципу классификации - *по количеству стратегий* - различают конечные и бесконечные игры. В *конечных* играх игроки располагают конечным числом возможных стратегий (например, в игре в орлянку игроки имеют по два возможных хода - они могут выбрать “орел” или “решку”). Сами стратегии в конечных играх нередко называются *чистыми* стратегиями (смысл этого названия будет ясен далее). Соответственно, в *бесконечных* играх игроки имеют бесконечное число возможных стратегий - так, в ситуации Продавец-Покупатель каждый из игроков может назвать любую устраивающую его цену и количество продаваемого (покупаемого) товара.

Третий способ классификации игр - по *свойствам функций* выигрыша (платежных функций). Важным случаем в теории игр является ситуация, когда выигрыш одного из игроков равен проигрышу другого, т.е. налицо прямой конфликт между игроками. Подобные игры называются *играми с нулевой суммой*, или *антагонистическими* играми. Игры в орлянку или в очко - типичные примеры антагонистических игр. Прямой противоположностью играм такого типа являются игры *с постоянной разностью*, в которых игроки и выигрывают, и проигрывают одновременно, так что им выгодно действовать сообща. Между этими крайними случаями имеется множество игр *с ненулевой суммой*, где имеются и конфликты, и согласованные действия игроков.

В зависимости от *возможности предварительных переговоров* между игроками различают кооперативные и некооперативные игры. Игра называется *кооперативной*, если до начала игры игроки образуют коалиции и принимают взаимообязывающие соглашения о своих стратегиях. Игра, в которой игроки не могут координировать свои стратегии подобным образом, называется *некооперативной*. Очевидно, что все антагонистические игры могут служить примером некооперативных игр. Примером кооперативной игры может служить ситуация образования коалиций в парламенте для принятия путем голосования решения, так или иначе затрагивающего интересы участников голосования.

13.3. Формальное представление игр

Дадим формальное описание перечисленных элементов конфликта. Множество всех игроков, обозначаемое I, в случае конечного их числа может задаваться простым перечислением игроков. Например, $I=\{1, 2\}$ при игре в орлянку, $I=\{$Продавец, Покупатель$\}$ в ситуации монополия-монопсония, $I=\{1, 2, ..., n\}$ в случае анализа результатов голосования в парламенте.

Множество стратегий игрока i обозначим через X_i. При игре в орлянку каждый игрок располагает двумя стратегиями: X_i={Орел, Решка}; каждый участник голосования имеет выбор на множестве стратегий {За, Против}. В случае взаимодействия на рынке как Продавец, так и Покупатель могут назначать некоторую неотрицательную цену на продаваемый (покупаемый) товар, т.е. множество стратегий каждого из них X_i: $P_i > 0$.

В каждой партии игрок выбирает некоторую свою стратегию x_i X_i, в результате чего складывается набор стратегий $x = \{x_1, x_2,... x_n\}$, называемый *ситуацией*. Так, ситуацию в Парламенте описывает список {За, За, Против, За ...}, полученный в итоге проведенного голосования.

Заинтересованность игроков в ситуациях проявляется в том, что каждому игроку i в каждой ситуации x приписывается число, выражающее степень удовлетворения его интересов в данной ситуации. Это число называется *выигрышем* игрока i и обозначается через $h_i(x)$, а соответствие между набором ситуаций и выигрышем игрока i называется *функцией выигрыша (платежной функцией)* этого игрока H_i.

В случае конечной игры двух лиц функции выигрыша каждого из игроков удобно представлять в виде *матрицы выигрышей*, где строки представляют стратегии одного игрока, столбцы - стратегии другого игрока, а в клетках матрицы указываются выигрыши каждого из игроков в каждой из образующихся ситуаций. (Данная форма представления конечных игр двух лиц объясняет общее для них название - *матричные игры*.)

Например, в случае игры в орлянку каждый из игроков имеет по две стратегии, именуемые Орел и Решка. Если игроки выбирают одинаковые стратегии, т.е. в случаях, если оба говорят “Орел” или оба говорят “Решка”, 1-й игрок выигрывает 1 рубль, а второй игрок проигрывает 1 рубль. В ситуациях, когда оба игрока выбирают различные стратегии, 1-й игрок проигрывает 1 рубль, а 2-й игрок соответственно этот 1 рубль выигрывает.

В итоге матрица выигрышей 1-го игрока H_1 выглядит следующим образом:

$$\begin{array}{lll} & & \text{Стратегии 2-го игрока} \\ & & \quad\text{Орел}\quad\text{Решка} \\ \text{Стратегии 1-го игрока} & \begin{array}{l}\text{Орел}\\ \text{Решка}\end{array} & \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array}$$

Соответственно матрица выигрышей 2-го игрока H_2 имеет вид:

$$\begin{array}{lll} & & \text{Стратегии 2-го игрока} \\ & & \begin{array}{cc}\text{Орел} & \text{Решка}\end{array} \\ \text{Стратегии 1-го игрока} & \begin{array}{l}\text{Орел} \\ \text{Решка}\end{array} & \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}$$

Для антагонистических игр, в которых выигрыш одного игрока равен проигрышу другого (игр с нулевой суммой), выполняется соотношение $H_1=-H_2$. Игра в орлянку, очевидно, является примером такой игры.

Часто для наглядности матрицы выигрышей для обоих игроков совмещают в одну, которая дает полное представление о всей игре:

$$\begin{array}{lll} & & \text{Стратегии 2-го игрока} \\ & & \begin{array}{cc}\text{Орел} & \text{Решка}\end{array} \\ \text{Стратегии 1-го игрока} & \begin{array}{l}\text{Орел} \\ \text{Решка}\end{array} & \begin{pmatrix} (1,-1) & (-1,1) \\ (-1,1) & (1,-1) \end{pmatrix} \end{array}$$

В каждой клетке этой матрицы слева указаны значения выигрыша 1-го игрока, справа - значения выигрыша 2-го игрока.

Рассмотрим пример задания матрицы выигрышей для игры с ненулевой суммой, называемой в литературе по теории игр *Дилемма Заключенного*. Содержание игры следующее: два преступника ожидают приговора суда за совершенное злодеяние. Адвокат конфиденциально предлагает каждому из преступников облегчить его участь (и даже освободить!), если он сознается и даст показания против сообщника, которому грозит угодить в тюрьму за совершенное преступление на 10 лет. Если никто не сознается, то обоим угрожает заключение на определенный срок (скажем, 1 год) по обвинению в незначительном преступлении. Если сознаются оба преступника, то, с учетом чистосердечного признания, им обоим грозит попасть в тюрьму на 5 лет. Каждый заключенный имеет на выбор 2 стратегии: не сознаваться или сознаваться, выдав при этом сообщника. В итоге можно получить следующую матрицу "выигрышей" для обоих игроков:

$$\begin{array}{lll} & & \text{Стратегии 2-го игрока} \\ & & \begin{array}{cc}\text{Сознаться} & \text{Не сознаться}\end{array} \\ \text{Стратегии 1-го игрока} & \begin{array}{l}\text{Сознаться} \\ \text{Не сознаться}\end{array} & \begin{pmatrix} (5,5) & (0,10) \\ (10,0) & (1,1) \end{pmatrix} \end{array}$$

Приведем, наконец, пример записи функции выигрыша для бесконечной игры. В случае дуополии каждый из игроков может объявить цену p_i, по которой он хотел бы продать некоторое количество товара. При этом предполагается, что потребители приобретут товар у фирмы, объявившей меньшую цену, или распределят свой спрос поровну между фирмами в случае, если они назначили одинаковую цену. Если функцию спроса в зависимости от цены на товар обозначить как $d(p)$, то функция выигрыша 1-й фирмы $\Pi_1(p_1, p_2)$ будет иметь вид

$$\Pi_1(p_1,p_2) = \begin{cases} p_1 d(p_1), \text{ если } p_1 < p_2, \\ p_1 \dfrac{d(p_1)}{2}, \text{ если } p_1 = p_2, \\ 0, \qquad \text{ если } p_1 > p_2. \end{cases}$$

Аналогично выглядит функция выигрыша 2-й фирмы $\Pi_2(p_1,p_2)$.

13.4. Принципы решения матричных антагонистических игр

В качестве основного допущения в теории игр предполагается, что каждый игрок стремится обеспечить себе максимально возможный выигрыш при любых действиях партнера. Предположим, что имеется конечная антагонистическая игра с матрицей выигрышей 1-го игрока H и, соответственно, матрицей выигрыша 2-го игрока $-H$. Пусть Игрок 1 считает, что какую бы стратегию он ни выбрал, Игрок 2 выберет стратегию, максимизирующую его выигрыш, и тем самым минимизирующую выигрыш Игрока 1. Оптимальная стратегия Игрока 1, которая обеспечит ему наибольший из возможных выигрышей вне зависимости от стратегии противника, будет состоять в выборе стратегии с самым высоким из таких платежей. Таким образом, Игрок 1 выбирает i-ю стратегию, которая является решением задачи

$$\max_i \min_j h_{ij}.$$

Игрок 2 точно так же стремится обеспечить себе наивысшую величину выигрыша (или, что эквивалентно, наименьшую величину проигрыша) вне зависимости от стратегии, избранной партнером. Его оптимальной стратегией будет столбец матрицы H с наименьшим значением максимального платежа. Таким образом, Игрок 2 выберет j-ю стратегию, которая является решением задачи

$$\min_j \max_i h_{ij}$$

В итоге, если Игрок 1 придерживается избранной стратегии (называемой *максиминной* стратегией), его выигрыш в любом случае будет не меньше максиминного значения (называемого также *нижней ценой игры*), т.е.

$$h_{ij} \geq \max_i \min_j h_{ij}$$

Соответственно, если Игрок 2 придерживается своей *минимаксной* стратегии, его проигрыш будет не больше минимаксного значения (называемого *верхней ценой игры*), т.е.

$$h_{ij} \leq \min_j \max_i h_{ij}$$

В случае, если верхняя цена игры равна нижней, т.е.

$$\max_i \min_j h_{ij} = \min_j \max_i h_{ij} = h_{ij}^*$$

оба игрока получают свои гарантированные платежи, а значение h_{ij}^* называется *ценой игры.* Элемент h_{ij} матрицы выигрышей, соответствующий максиминной и минимаксной стратегиям, называется *седловой точкой* матрицы H (это объясняется формой графика функции выигрыша в точке h_{ij}: она напоминает седло, убывая при изменении одной из переменных и возрастая при изменении другой переменной). В случае, если цена антагонистической игры равна 0, игра называется *справедливой.*

Рассмотрим игру, в которой Игрок 1 располагает двумя стратегиями, а Игрок 2 - тремя. Пусть матрица выигрышей Игрока 1 имеет следующий вид:

$$H_1 = \begin{bmatrix} 2 & 1 & 4 \\ -1 & 0 & 6 \end{bmatrix}.$$

(Поскольку мы рассматриваем пример антагонистической игры, матрица выигрышей Игрока 2 будет равна H_1, взятой с обратным знаком, т.е. $H_2 = - H_1$.)

Игрок 1 рассчитывает, что если он выберет первую стратегию (т.е. первую строку матрицы H_1), то противник может выбрать свою вторую стратегию (второй столбец), так что выигрыш будет равен 1. Если же он выберет вторую стратегию, то противник может избрать первую стратегию, так что выигрыш будет равен -1. Проанализировав полученные значения, Игрок 1 останавливается на своей первой стратегии, которая обеспечивает ему максимальный гарантированный выигрыш, равный 1.

Точно так же Игрок 2 рассматривает свои наихудшие варианты, когда противник выбирает свою первую стратегию, если Игрок 2 выбирает первую или вторую стратегии, или когда противник выбирает вторую стратегию, когда Игроком 2 выбран третий столбец. Эти варианты соответствуют максимальным значениям столбцов 2, 1 и 6. Взяв минимальное значение этих максимумов, Игрок 2 останавливается на своей второй стратегии, при которой его проигрыш минимален и равен 1:

$$\begin{array}{rccl}
 & & \min\limits_{j} h_{ij} & \\
 & \begin{bmatrix} 2 & 1 & 4 \\ -1 & 0 & 6 \end{bmatrix} & \begin{bmatrix} \underline{1} \\ -1 \end{bmatrix} & \leftarrow \max\limits_{i}\min\limits_{j} h_{ij} \\
\max\limits_{i} h_{ij} & (\;2\;\; \underline{1}\;\; 6\;) & & \\
 & \uparrow & & \\
 & \min\limits_{i}\max\limits_{j} h_{ij} & &
\end{array}$$

Следовательно, в этой игре существуют совместимые выборы, т.е.

$$h^{*} = \max_{i}\min_{j} h_{ij} = \min_{j}\max_{i} h_{ij} = 1.$$

В итоге будет разумно ожидать, что в описанной выше игре противники будут придерживаться избранных стратегий. Матричная антагонистическая игра, для которой $\max\limits_{i}\min\limits_{j} h_{ij} = \min\limits_{j}\max\limits_{i} h_{ij}$, называется *вполне определенной*, или игрой, *имеющей решение в чистых стратегиях*.

Однако далеко не все матричные антагонистические игры являются вполне определенными, и в общем случае

Игры, в которых выполняется строгое неравенство, называются *не полностью определенными* играми (или *не имеющими решения в чистых стратегиях* играми). Следующая матрица представляет пример подобной игры:

$$\begin{bmatrix} 6 & -2 & 3 \\ -4 & 5 & 4 \end{bmatrix}.$$

Для этой игры $\max\limits_{i}\min\limits_{j} h_{ij} = -2 < 4 = \min\limits_{j}\max\limits_{i} h_{ij}$. В результате, если игроки будут следовать предложенным выше правилам, то Игрок 1 выберет стратегию 1 и будет ожидать, что Игрок 2 выберет стратегию 2, при которой проигрыш равен -2, в то время как Игрок 2 изберет стратегию 3 и будет ожидать, что Игрок 1 выберет стратегию 2 с выигрышем, равным 4. Однако если Игрок 2 выберет свою третью стратегию, то Игрок 1 поступит правильнее, выбирая вторую, а не первую стратегию. Аналогично, если Игрок 1 выберет первую стратегию, то Игроку 2 выгоднее выбрать вторую стратегию, а не третью. По всей видимости, в играх такого типа принцип решения в чистых стратегиях оказывается непригодным.

В описанной ситуации игрокам становится важно, чтобы противник не угадал, какую стратегию он будет использовать. Для осуществления этого плана игрокам следует пользоваться так называемой *смешанной* стратегией. По существу, смешанная стратегия игрока представляет собой схему случайного выбора чистой стратегии. Математически ее можно представить как вероятностное распределение на множестве чистых стратегий данного игрока. В итоге вектор $x = (x_1, x_2, \dots x_m)$, где x_i 0 соответствует вероятности применения Игроком 1 i-й стратегии и $\sum_i x_i = 1$, задает смешанную стратегию этого игрока. Аналогично определяется смешанная стратегия Игрока 2 $y = (y_1, y_2, \dots y_n)$. Мы будем предполагать использование игроками их смешанных стратегий независимым, так что вероятность, с которой Игрок 1 выбирает i-тую стратегию, а Игрок 2 - j-ю, равна $x_i y_j$. В этом случае платеж равен h_{ij}. Суммируя по всем i и j, найдем математическое ожидание выигрыша Игрока 1:

$$v_1 = \sum_j \sum_i x_i h_{ij} y_j,$$

или в матричных обозначениях $v_1 = x\ H\ y'$.

На множестве своих смешанных стратегий Игрок 1, стремящийся достичь наибольшего из гарантированных выигрышей, выбирает вектор вероятностей x так, чтобы получить максимум минимальных значений ожидаемых выигрышей, т.е. решает задачу

$$v_1 = \max_x \min_y x\ H\ y'.$$

Аналогично целью Игрока 2 является достижение минимума максимальных значений своих проигрышей, т.е. он решает задачу

$$v_1 = \min_y \max_x x\ H\ y'.$$

Фундаментальным результатом теории игр является так называемая *Теорема о минимаксе*, которая утверждает, что сформулированные задачи для Игрока 1 и Игрока 2 всегда имеют решение для любой матрицы выигрышей H и, кроме того,

$$v_1 = v_2 = v.$$

Существующие доказательства этой теоремы основаны на теореме о неподвижной точке, или свойстве отделимости выпуклых множеств (см., например, Г.Н.Дюбин, В.Г.Суздаль. *Введение в прикладную теорию игр*).

Как и для вполне определенных игр, стратегия x^* Игрока 1 называется ***максиминной стратегией***, стратегия Игрока 2 y^* - ***минимаксной стратегией***, значение v - ***ценой игры***; в случае, когда $v = 0$, игра называется ***справедливой***.

Очевидным следствием из Теоремы о минимаксе является соотношение

$$x\ H\ y^* < \bar{V} < x^*\ H\ y,$$

которое означает, что никакая стратегия Игрока 1 не позволит выиграть ему сумму большую, чем цена игры, если Игрок 2 применяет свою минимаксную стратегию, и никакая стратегия Игрока 2 не даст возможности проиграть ему сумму меньшую, чем цена игры, если Игрок 1 применяет свою максиминную стратегию. Это верно также и для чистых стратегий как для частного случая смешанных стратегий (т.е. чистая стратегия - это стратегия, используемая с вероятностью 1): использование любой чистой стратегии в случае, если противник использует свою оптимальную стратегию, не позволяет выиграть больше (проиграть меньше) цены игры. Этот факт нередко используется для разработки конкретных алгоритмов решения антагонистических матричных игр.

13.5. Решение матричных антагонистических игр

Вообще говоря, оптимальные стратегии легко находятся для небольших игр, но вычисления становятся достаточно сложными с ростом числа стратегий. Для поиска оптимальных стратегий рекомендуется несколько подходов.

Для уменьшения размерности игры используется *доминирование* строк и столбцов. Обычно говорят, что k-я строка матрицы H доминирует i-тую строку (т.е. одна чистая стратегия доминирует другую), если

$$h_{ij} \le h_{kj} \text{ для всех } j,$$

$$h_{ij} < h_{kj} \text{ хотя бы для одного } j.$$

Аналогично l-й столбец доминирует j-й столбец, если

$$h_{il} \le h_{ij} \text{ для всех } i,$$

$$h_{il} < h_{ij} \text{ хотя бы для одного } i.$$

Смысл этого определения состоит в том, что доминирующая стратегия никогда не хуже, а в некоторых случаях даже лучше, чем доминируемая стратегия. Отсюда следует, что игроку нет необходимости использовать доминируемую стратегию. В самом деле, будут существовать оптимальные смешанные стратегии, при которых вероятность использования доминируемых строк и столбцов равна нулю, и при решении игры все доминируемые строки и столбцы могут быть отброшены, что позволяет уменьшить размеры матрицы. (Этот подход может использоваться также при поиске решения игры в чистых стратегиях.)

Рассмотрим, например, игру со следующей матрицей:

$$\begin{vmatrix} 5 & 2 & 1 \\ 2 & 1 & 3 \\ 3 & 6 & 4 \end{vmatrix}.$$

Третья строка этой матрицы доминирует вторую. Исключение второй строки приводит к матрице

$$\begin{vmatrix} 5 & 2 & 1 \\ 3 & 6 & 4 \end{vmatrix}.$$

Третий столбец в этой урезанной матрице доминирует второй, и исключение второго столбца дает

$$\begin{vmatrix} 5 & 1 \\ 3 & 4 \end{vmatrix}.$$

В итоге, если можно найти решение для полученной игры, то его легко использовать для решения исходной игры, просто приписав исключенным строкам и столбцам нулевые вероятности.

Другой метод упрощения матрицы выигрышей основывается на доказанном в теории игр свойстве, согласно которому *аффинное* преобразование матрицы платежей (т.е. преобразование всех элементов матрицы H по правилу $h'_{ij} = a\,h_{ij} + b$, где $a \neq 0$) не изменяет решения игры; кроме того, цена преобразованной игры v' может быть получена из цены первоначальной игры по такому же правилу: $v' = av + b$. Это означает, что для задания игры в принципе безразлично, в каких единицах измеряются выигрыши (например, в рублях или долларах); прибавление (вычитание) некоторой фиксированной суммы b изменит на такую же сумму выигрыш (проигрыш) каждого из игроков, не меняя решения игры.

Это свойство может быть использовано для упрощения и придания наглядности матрице выигрышей: так, если в клетках этой матрицы имеются дроби с общим знаменателем, всю матрицу можно умножить на некоторую константу для получения целых чисел; если большинство клеток матрицы заполнено одинаковыми элементами, их можно вычесть из всей матрицы для получения нулей в этих клетках. Кроме того, это свойство позволяет любую матричную антагонистическую игру сделать справедливой - для этого необходимо вычесть цену игры из всех элементов матрицы выигрышей.

Для решения игры 2×2 (и вообще игр 2×n или m×2) может быть использован, например, графический метод. Проиллюстрируем его на примере решения описанной игры в орлянку. Матрица выигрыша для этой игры, как было показано, имеет вид

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Пусть Игрок 1 избирает свою первую стратегию с вероятностью x и вторую стратегию с вероятностью $(1-x)$. Если Игрок 2 выбирает свою первую стратегию, то (из первого столбца матрицы) математическое ожидание выигрыша для Игрока 1 будет равно

$$v_1 = x - (1-x) = 2x - 1.$$

Если Игрок 2 выбирает свою вторую стратегию, то, в соответствии со вторым столбцом матрицы,

$$v_1 = -x + (1-x) = 1 - 2x.$$

Каждое из этих уравнений может быть изображено графически отрезком прямой линии в области [0,1] на графике с координатами x и v_1. Они представлены на левой части рис.1 соответственно как отрезки $A_1A'_1$ и $A_2A'_2$, которые пересекаются в точке P. При данном x на рис.1 показаны две величины v_1, которые Игрок 1 может получить, если Игрок 2 применяет свои чистые стратегии. Промежуточные значения v_1, соответствующие точкам между этими графиками, получаются, если Игрок 2 применяет смешанные стратегии. Меньшая из этих величин, соответствующая каждому значению x, показывает тот минимум, который может получить Игрок 1, выбирая стратегию $(x,(1-x))$. Следовательно, линия $A_1PA'_2$ показывает платеж, который Игрок 1 может гарантированно получить при любой стратегии Игрока 2. Игрок 1 выбирает такое значение x, чтобы достичь наивысшей точки. Для графика на рис. 13.1 этой точкой является точка P, для которой x=0,5 и v_1=0.

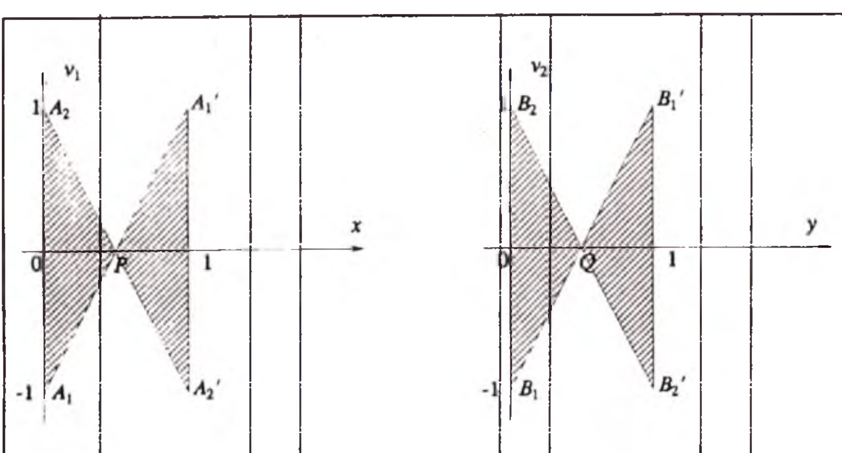


Рисунок 13.1. Пример графического решения матричной игры 2×2

Аналогично может быть проанализирована игра для Игрока 2, который использует свою первую стратегию с вероятностью y, а вторую - с вероятностью $(1-y)$. В правой части рис. 13.1 ломанная линия B_2QB_1 представляет наибольший проигрыш Игрока 2 при различных выборах y. Игрок 2 выбирает y так, чтобы достичь низшей точки на этой линии, т.е. точки Q, для которой $y = 0{,}5$ и $v_2 = 0$. Следовательно, игроки в орлянку должны применять свои стратегии с одинаковой вероятностью 0,5, и цена игры при этом будет равна нулю, т.е. описанная игра справедлива.

Другой метод решения матричных игр базируется на указанном выше свойстве, гарантирующем, что применение чистых стратегий Игроком 2 (Игроком 1) против оптимальной смешанной стратегии Игрока 1 не позволяет ему проиграть меньше (выиграть больше), чем значение цены игры v. Это позволяет сформулировать следующую задачу для Игрока 1:

$$\max v$$

$$\begin{cases} h_{11}x_1 + h_{21}x_2 + \dots + h_{m1}x_m \le v \\ h_{21}x_1 + h_{22}x_2 + \dots + h_{m2}x_m \le v \\ \dots \\ h_{1n}x_1 + h_{2n}x_2 + \dots + h_{mn}x_m \le v \\ x_1 + x_2 + \dots x_m = 1, \ x_i \ge 0. \end{cases}$$

Выписанная задача - типичная задача линейного программирования и легко поддается решению его методами. Задача для Игрока 2 выглядит аналогично и является двойственной задачей линейного программирования к задаче Игрока 1.

Таким образом, в общем случае для решения матричной антагонистической игры размерностью $m \times n$ необходимо решить пару двойственных задач линейного программирования, в результате чего находится набор оптимальных стратегий x^*, y^* и цена игры v.

13.6. Игры с ненулевой суммой и кооперативные игры

В игре с ненулевой суммой уже становится необязательно, чтобы один из участников выигрывал, а другой проигрывал; напротив, они могут и выигрывать, и проигрывать одновременно. Поскольку интересы игроков теперь не являются полностью противоположными, их поведение становится более разнообразным. Так, например, если в игре с нулевой суммой каждому игроку невыгодно было сообщать другому свою стратегию (это могло уменьшить его выигрыш), то в игре с ненулевой суммой становится желательным както координировать свои действия с партнером или каким-либо способом влиять на его действия.

Игры с ненулевой суммой могут быть кооперативными и некооперативными. В некооперативных играх игроки принимают решения независимо друг от друга либо потому, что осуществление соглашения невозможно, либо потому, что оно запрещено правилами игры. Описанная выше игра *Дилемма Заключенного* представляет пример игры двух лиц с ненулевой суммой, в которой взаимодействие игроков невозможно по условиям игры.

Один из подходов к решению некооперативных игр состоит в определении *точек равновесия* игры. Понятие равновесия в теории игр шире понятия оптимальности в теории оптимизации и включает последнее в качестве частного случая. В общем случае пара стратегий X, Y для Игрока 1 и Игрока 2 называется *точкой равновесия по Нэшу*, если ни одному из игроков невыгодно отклоняться от своей стратегии в одиночку, т.е. если

$$H_1(X,\ Y^*) \le\ H_1(X^*,\ Y^*) \text{ для любых } X \text{ и}$$
$$H_2(X^*,\ Y) \le\ H_2(X^*,\ Y^*) \text{ для любых } Y.$$

(Это определение равновесия сохраняется и для игры n лиц.)

Рассмотрим пример, когда матрица выигрышей игры имеет следующий вид:

$$\begin{vmatrix} (4,1) & (0,0) \\ (0,0) & (1,4) \end{vmatrix}$$

Легко видеть, что в данной игре пары стратегий $x = (1, 0)$, $y = (1, 0)$ и $x = (0, 1)$, $y = (0, 1)$ являются равновесными, т.е. Игроку 2 (1) не выгодно отклоняться от 1-й (2-й) стратегии, если Игрок 1 (2) придерживается 1-й (2-й) стратегии. Отметим также, что выигрыши в равновесных точках различны.

Доказано, что для любой конечной некооперативной игры с ненулевой суммой (называемой также *биматричной* игрой) всегда существует по крайней мере одна равновесная пара смешанных стратегий. В общем случае равновесное решение может быть неединственным, и каждому из них могут соответствовать различные значения выигрыша каждого из игроков.

Кооперативной игрой называется игра с ненулевой суммой, в которой игрокам разрешается обсуждать перед игрой свои стратегии и договариваться о совместных действиях, т.е. игроки могут образовывать коалиции. Основная задача в кооперативной игре состоит в дележе общего выигрыша между членами коалиции.

В случае игры двух лиц предполагается, что два игрока не могут воздействовать друг на друга, пока не придут к некоторому соглашению. Таким образом, игра определяется как множество S в пространстве переменных h_1 и h_2, представляющее общие выигрыши;

Рисунок 13.2. Множество возможных выигрышей, Парето-оптимальное множество, точка угрозы, переговорное множество и решение Нэша в кооперативной игре 2-х лиц

кроме того, заданы два числа T_1, T_2, определяющие величины выигрыша, которые каждый из игроков может получить, не вступая в коалицию со своим партнером. Обычно предполагают, что множество S является замкнутым, выпуклым и ограниченным сверху. Точка T с координатами (T_1, T_2) называется *точкой угрозы* (см. рис.13.2).

На множестве возможных выигрышей выделяется множество *Парето-оптимальных решений*, т.е. множество точек, принадлежащих S, для которых увеличение выигрыша одного из игроков возможно только за счет уменьшения выигрыша его партнера. Очевидно, множество таких точек образует северо-восточная граница множества S.

Все точки Парето-оптимального множества, находящиеся одновременно выше и правее точки угрозы T, образуют так называемое *переговорное множество*. Очевидно, что игрокам нет смысла договариваться относительно решений, не принадлежащих переговорному множеству, либо потому, что положение одного из игроков может быть улучшено при сохранении положения его партнера и можно договариваться о более выгодных решениях, либо потому, что по крайней мере для одного из игроков теряет смысл вступать в коалицию со своим партнером - не худших результатов он может достичь и в одиночку.

Наконец, на переговорном множестве выделяется *точка решения Нэша* N, в которой достигается максимум произведения превышения выигрышей каждого из игроков над платежами, которые могут быть получены без вступления в коалицию:

$$\max (h_1 - T_1)(h_2 - T_2).$$

В теории игр показано, что если множество возможных платежей S выпукло, замкнуто и ограничено сверху, то точка Нэша N существует и единственна. Точка Нэша представляет одно из возможных решений кооперативной игры, от которого нет оснований отказываться ни одному из игроков.

Проиллюстрируем введенные понятия на примере игры, известной под названием *Семейный спор*. Согласно условиям этой игры, семейная пара - Муж и Жена каждый вечер решают проблему: как им провести свой досуг. В городке, где они живут, имеется два вида развлечений: Балет и Футбол. У каждого из супругов есть свое любимое зрелище: Жена предпочитает Балет, Муж - Футбол. Однако супруги так привязаны друг к другу, что посещение любимого развлечения в одиночку доставляет им совсем не такое удовольствие, как присутствие на них вдвоем, т.е. если Жена идет вечером на Балет с Мужем, она получает максимум удовольствия (скажем, 4 единицы); Муж недолюбливает Балет, но присутствие на нем с Женой скрашивает тягостное времяпровождение (Муж получает 1 ед. удовольствия). История повторяется с точностью до наоборот, когда Жена идет с Мужем на обожаемый им Футбол: Муж получает 4 ед. удовольствия от игры любимой команды и присутствия любимой Жены; Жена получает 1 ед. удовольствия, проведя вечер с Мужем на Футболе. В принципе Муж может сходить на Футбол, а Жена - на Балет в одиночку, но отсутствие супруга снижает удовольствие от любимых зрелищ - каждый из них получает по 2 ед. удовольствия. И наконец, вечер будет проведен совсем уж без пользы (т.е. супруги получат по 0 ед. удовольствия), если Муж отправится на Балет в то время как Жена будет на стадионе смотреть Футбол.

В итоге матрица выигрышей описанной игры выглядит следующим образом:

$$\begin{array}{lll} & & \text{Жена} \\ & & \begin{array}{cc}\text{Балет} & \text{Футбол}\end{array} \\ \text{Муж} & \begin{array}{r}\text{Балет} \\ \text{Футбол}\end{array} & \begin{pmatrix} (4,1) & (0,0) \\ (2,2 & (1,4) \end{pmatrix}. \end{array}$$

Можно показать, что если супруги будут придерживаться различных несогласованных смешанных стратегий, множество возможных выигрышей образует в системе координат значений выигрышей супругов h_1, h_2 треугольник ABC с вершинами в точках (0, 0), (1, 4), (4, 1) (рис. 13.3). Линия BC является множеством Парето-

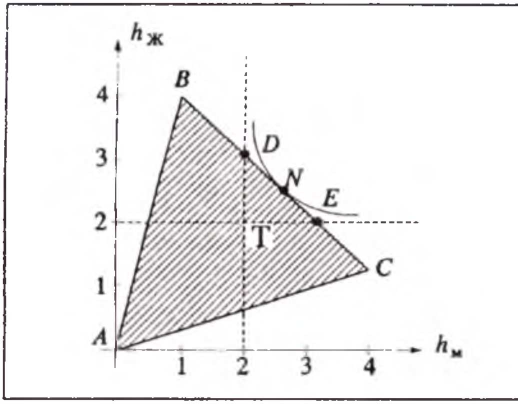


Рисунок 13.3. Решение кооперативной игры "Семейный спор".

оптимальных решений: вдоль этой линии рост удовольствия, получаемого Женой, возможен только за счет снижения удовольствия Мужа. Точка T с координатами (2, 2) является точкой угрозы в этой игре, а "угроза", например, со стороны Жены может звучать буквально следующим образом: "Вместо того, чтобы более 2/3 своего свободного времени проводить на этом Футболе, я буду ходить на Балет (с Мужем или без него - не важно) - ничего не потеряю." Аналогично может звучать "угроза" Мужа.

В итоге переговорное множество, образуемое точкой угрозы T, представлено линией DE на Парето-оптимальном множестве решений BC (рис. 13.3). На линии DE Муж и Жена могут договариваться, как часто они будут бывать вместе на одном из зрелищ; но при этом, во избежание взаимных угроз, ни одному из развлечений они не должны уделять более своих свободных вечеров.

Решение Нэша, когда максимально произведение приростов удовольствия Мужа и Жены по сравнению с удовольствием от независимого посещения Футбола и Балета, представлено точкой N на рис. 13.3 - супруги договариваются половину своего свободного времени проводить вместе на Балете, вторую половину - на Футболе.

13.7. Применение аппарата теории игр для анализа проблем микроэкономики

Введенные понятия математической теории игр имеют прямое приложение для анализа проблем микроэкономики, в т.ч. для анализа рыночного равновесия как кооперативной игры многих лиц.

Представим себе экономику, в которой имеется два субъекта: Игрок 1 и Игрок 2, и два товара (блага): x_1 и x_2. (Естественно, число игроков и товаров может быть большим, но в случае 2×2 все введенные понятия имеют наглядную интерпретацию.)

Каждый из игроков имеет свою функцию полезности, заданную на наборе товаров: $h_1(x_1,x_2)$, $h_2(x_1,x_2)$; предполагается, что эти функции непрерывны и монотонны по каждой из переменных и выпуклы. В начале игры в экономике имеется общее количество X_1 первого товара и X_2 - второго товара. Предположим, что это начальное количество благ как-то распределено между игроками: 1-й Игрок обладает количеством X^1_1 первого товара и X^1_2 - второго, 2-й Игрок - количествами X^2_1 и X^2_2 1-го и 2-го товаров соответственно, так что $X^1_1+X^2_1=X_1$ и $X^1_2+X^2_2=X_2$.

Встают вопросы: могут ли игроки путем обмена имеющимися у них товарами улучшить свое положение, т.е. увеличить значение функций полезности h_1 и h_2 по сравнению с начальными уровнями $h_1(X^1_1,X^1_2)$ и $h_2(X^2_1,X^2_2)$; каковы свойства такого решения?

Для наглядного представления экономики с двумя игроками и двумя товарами традиционно используется так называемый *ящик Эджворта* (рис. 13.4). В ящике Эджворта длина горизонтальной оси, соответствующей первому товару, равна общему количеству этого товара X_1, длина вертикальной оси - общему количеству товара X_2. Выделенное пространство является множеством всех возможных распределений имеющихся товаров между двумя игроками. Нижний левый угол считается началом координат для 1-го Игрока, верхний правый угол - началом координат для 2-го Игрока. На выделенном пространстве представлены также два множества кривых безразличия (линий уровня функций выигрыша), принадлежащих каждому из игроков. При этом точка начального распределения товаров имеет координаты (X^1_1,X^1_2) в системе отсчета 1-го Игрока (и, соответственно, (X^2_1,X^2_2) в системе отсчета 2-го Игрока).

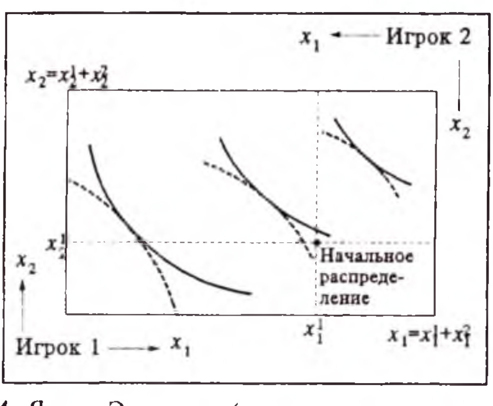


Рисунок 13.4. Ящик Эджворта (здесь и далее сплошными линиями обозначены кривые безразличия 1-го Игрока, пунктирными - кривые безразличия 2-го Игрока)

Рассмотрим для начала проблему эффективного распределения товаров между игроками. Единственным требованием к распределению, которое мы можем предъявить на начальном этапе анализа, является требование Парето-оптимальности. Напомним, что распределение называется Парето-оптимальным, если положение ни одного из игроков нельзя улучшить, не ухудшая при этом положение его партнера.

Множество Парето-оптимальных распределений может быть наглядно представлено с помощью ящика Эджворта. В случае 2-х игроков Парето-оптимальное решение может быть найдено с помощью фиксации уровня полезности одного из игроков (скажем, Игрока 2) и поиска максимума функции полезности другого игрока.

В терминах ящика Эджворта это означает, что необходимо найти такую точку на фиксированной кривой безразличия Игрока 2, в которой Игрок 1 получает максимум своей функции полезности (рис.5). Очевидно, что такой точкой является точка, где кривые безразличия касаются друг друга, так как в противном случае Игрок 1 может, продвигаясь вдоль фиксированной линии уровня Игрока 2 внутрь, увеличить значение своей функции полезности (рис. 13.5). Это свойство можно доказать и математически: максимум функции полезности Игрока 1 при фиксированном уровне полезности Игрока 2 достигается в точке, в которой дифференциалы этих функций равны, т.е. в точке, где кривые безразличия имеют общую касательную.

Опираясь на этот факт, можно показать, что множество Парето-оптимальных распределений в ящике Эджворта будет множеством всех точек, в которых кривые безразличия Игрока 1 и Игрока 2 касаются друг друга (рис. 13.6). Множество Парето-оптимальных

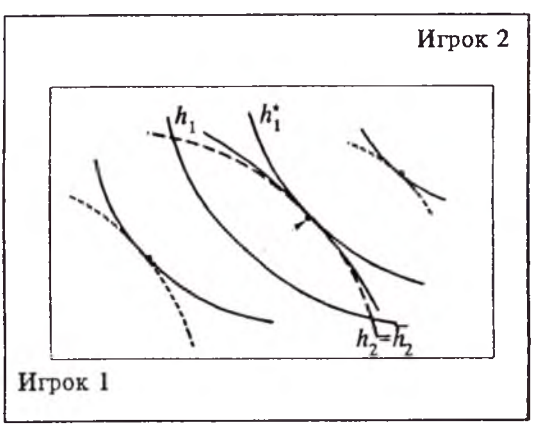


Рисунок 13.5. Парето-оптимальное распределение в ящике Эджворта

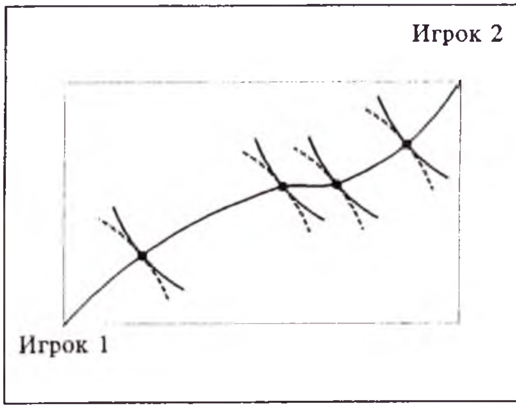


Рисунок 13.6. Множество Парето-оптимальных распределений (контрактное множество) для ящика Эджворта

распределений в пространстве товаров называется *контрактным* множеством, поскольку игрокам в общем случае имеет смысл договариваться между собой именно на этом наборе эффективных распределений (ср. с множеством Парето-оптимальных стратегий в кооперативных играх).

Рассмотрим теперь ситуацию, когда каждый игрок обладает некоторым начальным количеством каждого из товаров. Встает вопрос: может ли это начальное распределение быть улучшено путем обмена товарами между игроками? Исследуем эту проблему с помощью ящика Эджворта. Пусть (X^1_1, X^1_2) - точка начального распределения товаров; проведем через эту точку кривые безразличия для Игрока 1 и Игрока 2 (рис. 13.7).

Рисунок 13.7. Линии угрозы, переговорное множество и решение Нэша для ящика Эджворта

Если эти две кривые не касаются друг друга (т.е. если начальное распределение не является Парето-оптимальным), то в своем пересечении они образуют область, двигаясь внутрь которой каждый из игроков может увеличивать значение обеих функций полезности. При этом, как легко показать, часть контрактного множества оказывается *внутри* области, образованной проведенными кривыми безразличия.

Как мы уже убедились, игрокам имеет смысл вести переговоры относительно распределений, находящихся на контрактном множестве, а с учетом начального распределения - относительно участка контрактного множества, заключенного между двумя кривыми безразличия. Эти кривые, фиксируя начальные уровни полезности для каждого из игроков, аналогичны точке угрозы в теории кооперативных игр (поэтому назовем их *линиями угрозы*), а выделяемый ими участок на контрактном множестве - переговорному множеству (в микроэкономике оно также называется *переговорным* множеством). Для того чтобы переместиться на переговорное множество, в случае рис. 13.7 Игрок 1 должен передать (продать) некоторое количество имеющегося у него товара 1 Игроку 2 в обмен на определенное количество товара 2, имеющегося у Игрока 2.

Линии угрозы в данном случае означают, что за их пределами (т.е. ниже и левее исходной кривой безразличия для Игрока 1 и выше и правее кривой безразличия для Игрока 2) какому-либо из игроков становится незачем вести переговоры - ему лучше (или, по крайней мере, не хуже) оставаться в ситуации начального распределения. Точка Нэша, соответствующая максимуму произведения приращений полезностей игроков по сравнению с начальной ситуацией, будет находиться внутри переговорного множества.

В результате проведенного анализа можно сделать вывод, что игроки могут улучшить свое первоначальное положение, обмениваясь товарами, и Игроку 1 выгодно уступить Игроку 2 некоторое количество товара 1 в обмен на товар 2.

Рассмотрим в заключение решение сформулированной в разделе 3 задачи о дуополии. В этой задаче, как было сказано, две фирмы сталкиваются с проблемой удовлетворения спроса на некоторый товар. Объем спроса зависит от уровня назначаемых цен и описывается функцией $d(p)$ (ей соответствует нисходящая линия на рис. 13.8). Объем предложения товара каждой из фирм также зависит от уровня цен и в микроэкономике описывается функциями предложения $s_1(p)$, $s_2(p)$; эти функции определяются уровнем предельных издержек каждой из фирм. Предположим для простоты, что Фирма 1 и Фирма 2 имеют одинаковые функции предложения $s_1(p)=s_2(p)$ (рис. 13.8).

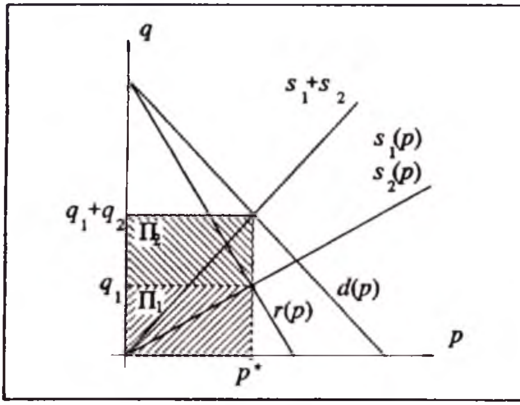


Рисунок 13.8. Решение задачи о дуополии

Поиск решения в задаче о дуополии (т.е. определение уровня цен и объемов предложения каждой из фирм) базируется на принципах, общих для решения задач теории игр: каждая из сторон располагает информацией о себе и своем партнере (в данном случае - о функциях предложения каждой из фирм), об условиях игры (в данном случае - о функции спроса) и действует, исходя из предположения, что ее партнер располагает такой же информацией и действует рационально (т.е. стремится максимизировать свой доход).

Если Фирма 1 назначит цену на предлагаемый ею товар p_1, а Фирма 2 примет эту цену, то Фирма 1 сможет продать объем товара, равный

$$r_1(p_1) = d(p_1) - s_2(p_1).$$

Функция $r_1(p)$ называется *остаточной функцией спроса*, с которой сталкивается Фирма 1 (рис. 13.8). Поскольку величина r_1 описывает объем спроса, приходящийся только на продукцию Фирмы 1, то она получит максимум дохода, полностью удовлетворив этот спрос, т.е. при условии, что

$$s_1(p_1) = r_1(p_1).$$

В итоге Фирма 1, опираясь на имеющуюся у нее информацию, решает задачу поиска равновесного уровня цен p , при которых

$$s_1(p^*) = d(p^*) - s_2(p^*).$$

Аналогичную задачу поиска равновесных цен решает Фирма 2:

$$s_2(p^*) = d(p^*) - s_1(p^*).$$

Учитывая, что $s_1(p)=s_2(p)$, мы получим, что в ситуации равновесия

$$s_1(p^*) = s_2(p^*) = \frac{d(p^*)}{2},$$

а доход каждой из фирм будет равен

$$П_1(p^*) = П_2(p^*) = p^* \frac{d(p^*)}{2}.$$

Таким образом, в задаче о дуополии фирмы должны найти такой уровень цен p^*, при котором они смогут полностью удовлетворить спрос на продукцию $d(p^*)$, распределив между собой производство этой продукции поровну и получив при этом одинаковый доход. Уровень равновесных цен и объем предложения каждой из фирм определяют в данной задаче ситуацию равновесия по Нэшу (рис. 13.8).

В итоге, как мы видим, проблемы рыночного взаимодействия близки к проблемам теории игр и могут быть эффективно описаны и исследованы в ее терминах.

13.8. Позиционные игры

Рассмотрим в заключение еще один пример анализа рыночного поведения с помощью аппарата теории игр, когда задача по своей структуре несколько отличается от задач, обсуждавшихся ранее.

Предположим следующую ситуацию. На рынке некоторого продукта доминирует производитель-монополист (Фирма 1), и монопольное положение приносит ему 12 млрд.руб. прибыли. Высокая прибыль в данном секторе привлекает других производителей, и, в частности, Фирма 2 решает вопрос: построить ли ей свой завод и начать на нем производство такого же товара? Однако ей известно, что Фирма 1 может предпринять некоторые действия в ответ на вторжение. С одной стороны, Фирма 1 может снизить объем своего производства, уступая часть рынка Фирме 2 и деля с ней получаемую прибыль - так, как это происходило в примере поведения фирм-олигополистов. В этом случае каждая из фирм получит по 6 млрд.руб. прибыли. С другой стороны, Фирма 1 может сохранить объем своего производства. В этом случае рост совокупного предложения товара Фирмами 1 и 2 снизит цену на этот товар, и, как следствие, прибыль Фирмы 1 упадет до 5 млрд.руб. Одновременно снижение цен приведет к тому, что Фирма 2, сделавшая предварительные затраты для выхода на новый для нее рынок, понесет чистые убытки: она потеряет на этом деле 2 млрд.руб. В случае, если Фирма 2 воздерживается от вступления на рынок, она ничего не выигрывает и не проигрывает (ее прибыль равна 0 млрд.руб.), а Фирма 1 продолжает получать монопольную прибыль в 12 млрд.руб. Если же Фирма 1 вдруг решит в этой ситуации снизить объем своего производства, ее прибыль упадет до 8 млрд.руб.

В принципе сформулированная конечная неантагонистическая игра двух лиц может быть описана следующей матрицей выигрышей (первыми указаны выигрыши Фирмы 1 в млрд.руб.):

Стратегия Фирмы 1

Сохранить объем производства — Снизить объем производства

$$\begin{pmatrix} (5,-2) & (6,6) \\ (12,0) & (8,0) \end{pmatrix}$$

Однако заметим, что описанная игра по своим условиям отличается от уже рассмотренных игр. Если ранее мы предполагали, что игроки принимают свои решения одновременно, не зная о решении партнера (что было весьма существенно!), то в данной игре Фирма 1 принимает решение, уже зная о решении, избранном Фирмой 2, в ответ она действия Фирмы 2, и это в корне меняет ситуацию.

Игры подобного типа, где задается последовательность принятия решений игроками, называются *позиционными играми*; число игроков и шагов в них может равняться 2 (как в нашем примере), 3 и т.д. К позиционным многошаговым играм двух лиц, где игроки принимают решения, зная о всех предыдущих решениях партнера, можно отнести, например, шахматы и шашки.

В силу отмеченных особенностей структуры позиционной игры ее более наглядно представляет не матрица выигрышей, а *дерево решений* (или, в общем случае, *граф решений*), приводящее игроков из исходной позиции в конечные. Так, описанную игру *Вступление на рынок* можно представить следующим деревом (рис. 13.9), ветви которого соответствуют решениям партнеров, а у каждой из висячих вершин указаны выигрыши игроков (как и ранее, первыми указаны выигрыши Фирмы 1, в млрд.руб.).

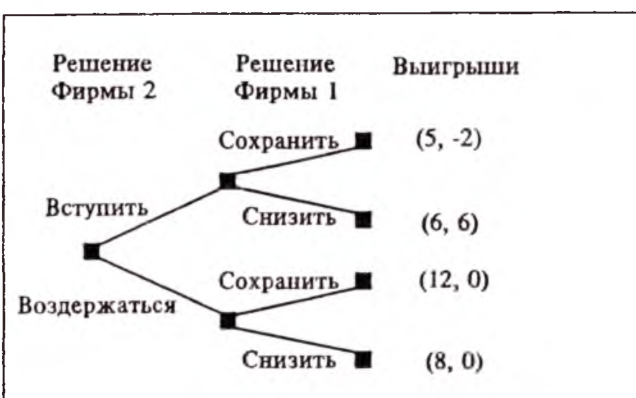


Рисунок 13.9. Дерево решений для игры *Вступление на рынок*

Вершины дерева игры называются *позициями*; позиции, непосредственно следующие за некоторой позицией, называются *альтернативами*; позиции, не имеющие альтернатив, называются *окончательными*, а ведущие в них пути - *партиями* (так, описанная игра имеет четыре партии). Часть дерева решений, описывающая игру из некоторой позиции после нескольких начальных шагов партнеров, называется *подыгрой*, и ее решение может представлять самостоятельную задачу. (Хорошим примером подыгр являются шахматные этюды типа "За сколько шагов из данной позиции белые смогут поставить мат черному королю?")

Описанная игра *Вступление на рынок* имеет две пары стратегий (две партии), удовлетворяющих условию равновесия по Нэшу: партия, когда Фирма 2 решает воздержаться от вступления на рынок, а Фирма 1 сохраняет объем своего производства, и партия, когда Фирма 2 решает вступить на рынок, а Фирма 1, в свою очередь, снижает объем производства. Легко убедиться, что в каждой из этих двух партий отступление каждого из игроков от своей стратегии приводит к уменьшению его выигрыша.

Возникает вопрос: реализация какой из этих двух равновесных партий наиболее вероятна? В непозиционной игре, в которой игроки принимают решение одновременно и независимо друг от друга, реализация обеих партий была бы равновероятна, т.е. у исследователя нет никаких причин ожидать, что один из исходов будет встречаться чаще при многократной реализации этой игры.

Однако в позиционной игре необходимо учитывать, что Фирма 1 принимает решение, уже зная о решении, принятом Фирмой 2. При этом менеджеры Фирмы 2, которая должна сделать первый шаг, при выборе своей стратегии могут рассуждать следующим образом: "Если мы не вступим на рынок со своей продукцией, то в любом случае мы ничего не потеряем. С другой стороны, если мы решим внедриться на рынок, не исключено, что Фирма 1 сохранит объем своего производства, и для нас это обернется потерями в 2 млрд.руб.!" Затем, следуя принципу максимизации своего минимального выигрыша, Фирма 2 должна была бы избрать стратегию "Воздержаться от вступления на рынок" - ее прибыль в этом случае максимальна (0 млрд.руб. больше, чем -2 млрд.руб.).

Эти, казалось бы, логичные рассуждения не учитывают одной из главных предпосылок теории игр - предположения о рациональном поведении игроков, стремящихся к максимизации своих выигрышей. В данном случае это заставляет менеджеров Фирмы 2 задать себе вопрос: "А насколько вероятна реализация Фирмой 1 стратегии "Сохранить объем производства", если мы вторгнемся на рынок?

Ведь в этом случае Фирма 1 получит меньшую прибыль (5 млрд.руб.), чем в случае, если она снизит объем своего производства и поделится частью рынка с нами, получив при этом 6 млрд.руб." В итоге, учитывая, что Фирма 1 будет вести себя рационально, ее ответом на вступление Фирмы 2 на рынок должно стать снижение объема своего производства, а не реализация угрозы сохранить прежний объем производства и подавить Фирму 2.

В данном случае в теории игр речь идет о *правдоподобности* угроз. В обсуждаемой игре угроза Фирмы 1 сохранить объем производства в ответ на вторжение Фирмы 2 на рынок является *неправдоподобной*, поскольку ее реализация приводит к меньшему выигрышу по сравнению с другими исходами. Учитывая этот факт, можно утверждать, что наиболее вероятной будет реализация партии, когда Фирма 2 вступает на рынок, а Фирма 1 в ответ на это вторжение снижает объем своего производства: эта партия равновесна по Нэшу и, кроме того, учитывает степень правдоподобности угрозы Фирмы 1 сохранить объем своего производства и подавить таким образом Фирму 2.

Рассмотрим пример игры *Вступление на рынок* с несколько измененными исходными данными. Дерево решений этой игры и выигрыши фирм в млрд.руб. представлены на рис. 13.10.

Анализ игры, представленной на рис. 13.10, показывает, что угроза Фирмы 1 сохранить объем своего производства является вполне правдоподобной: прибыль Фирмы 1 в этом случае (4 млрд.руб) не меньше, чем в случае, если она снизит объем своего производства и уступит часть рынка Фирме 2 (3 млрд.руб.). И вообще, стратегия Фирмы 1 "Сохранить объем производства" является до-

Рисунок 13.10. Пример дерева решений для игры *Вступление на рынок*

минирующей, т.е., следуя этой стратегии, Фирма 1 получает большую прибыль, чем в случае реализации другой своей стратегии "Снизить объем производства", - независимо от решений Фирмы 2. Учитывая этот факт, Фирма 2, во избежание лишних потерь, должна избрать стратегию "Воздержаться от вступления на рынок". Как следствие, в данном случае наиболее вероятной является реализация равновесной партии, когда Фирма 2 воздерживается от вступления на рынок, а Фирма 1 сохраняет объем своего производства, продолжая пользоваться монопольным положением.

Рассмотренный пример описывает случай так называемой *устойчивой* монополии, когда фирма-монополист в состоянии эффективно реализовать угрозу подавления своих потенциальных партнеров. Это может объясняться естественными условиями производства, его технологическими особенностями, факторами, позволяющими фирме-монополисту гибко реагировать на действия противников (большой запас мощностей, реклама, вложения в перспективные исследования и т.п.). Поэтому вполне естественным выглядит желание государственных органов контролировать, а по возможности и ограничивать деятельность подобных фирм - устойчивых монополистов.

В итоге, как мы видим, проблемы рыночного взаимодействия близки к проблемам теории игр и могут быть эффективно описаны и исследованы в ее терминах.

Вопросы к главе 13

1. Чем отличаются проблемы теории игр от проблем теории оптимизации?
2. Какие встречаются типы игр?
3. Как определяется матричная антагонистическая игра двух лиц?
4. Как находится верхняя и нижняя цена игры для вполне определенной матричной антагонистической игры двух лиц?
5. Всегда ли матричные игры имеют решение в чистых стратегиях? Каковы принципы решения не вполне определенных матричных игр?
6. Сформулируйте Теорему о минимаксе.
7. Какие есть методы упрощения и решения матричных антагонистических игр?
8. В чем отличие игр с ненулевой суммой от антагонистических игр? Чем отличаются кооперативные игры от некооперативных?
9. Дайте определение Парето-оптимального множества, переговорного множества и решения Нэша для кооперативных игр.
10. Какова связь между проблемами теории игр и микроэкономики?
11. В чем особенность позиционных игр? Как степень правдоподобности угроз одного из партнеров влияет на исход позиционной игры?

ГЛАВА 14
ЭКОНОМИЧЕСКИЕ МОДЕЛИ И СТАТИСТИЧЕСКИЕ МЕТОДЫ. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Целью этой и последующих глав является ознакомление читателя с методами исследования (проверки, обоснования, оценивания) количественных закономерностей и качественных утверждений (гипотез) в экономике на основе анализа статистических данных. Эти методы являются составной частью эконометрики - науки, изучающей экономические явления с количественной точки зрения. Эконометрика устанавливает и исследует количественные закономерности в экономике на основе методов теории вероятности и математической статистики, адаптированных к обработке экономических данных.

Закономерности в экономике выражаются в виде связей и зависимостей экономических показателей, математических моделей их поведения. Такие зависимости и модели могут быть получены только путем обработки реальных статистических данных, с учетом внутренних механизмов связи и случайных факторов. Модель может быть получена и апробирована на основе анализа статистических данных, и изменения в поведении последних говорят о необходимости уточнения и развития модели. Особенно важен эконометрический анализ в макроэкономике, где взаимосвязи величин зачастую неочевидны и изменчивы. Нередко встречается ситуация, когда модель перестает "работать" в связи с появлением или активизацией какого-то фактора, и такие ситуации обусловливают развитие макроэкономической теории. Поэтому предлагаемый материал "привязан" к макроэкономическим проблемам и моделям. Эконометрический анализ дает возможность обосновать и уточнить форму зависимостей в рассматриваемых макроэкономических моделях, лучше понять механизмы взаимосвязи макроэкономических показателей.

Основным элементом экономического исследования является анализ и построение *взаимосвязей экономических переменных*. Изучение таких взаимосвязей осложнено тем, что они, особенно - в макроэкономике, не являются строгими, функциональными зависимостями. Во-первых, всегда очень трудно выявить все основные факторы, влияющие на данную переменную. Во-вторых, многие такие воздействия являются случайными, то есть содержат случайную составляющую. В-третьих, экономисты, как правило, располагают ограниченным набором данных статистических наблюдений, которые к тому же содержат различного рода ошибки. *Математическая статистика* (то есть теория обработки и анализа данных) и ее применение в экономике - *эконометрика* - позволяют строить экономические модели и оценивать их параметры, проверять гипотезы о свойствах экономических показателей и формах их связи, что в конечном счете служит основой для экономического анализа и прогнозирования, создавая возможность для принятия обоснованных экономических решений.

Любое эконометрическое исследование всегда предполагает объединение теории (экономической модели) и практики (статистических данных). Мы используем теоретические модели для описания и объяснения наблюдаемых процессов и собираем статистические данные с целью эмпирического построения и обоснования моделей.

14.1. Введение случайного компонента в экономическую модель

Обычно предполагают, что все факторы, не учтенные явно в экономической модели, оказывают на объект некоторое результирующее воздействие, величина которого неизвестна заранее и может быть описана как случайная функция. Для ее описания в модель добавляют (обычно аддитивным образом) случайный параметр ε, интегрирующий в себе влияние всех неучтенных явно факторов. Например, в модели спроса

$$q = f(p,\ I) + \varepsilon \tag{1}$$

(q - количество блага, p - цена, I - доход потребителя) переменная ε учитывает влияние всех прочих факторов (цен на другие товары, изменений моды, погоды и т. д.), не учтенных явно в функции спроса.

14.2. Статистические данные и стохастическая модель. Эконометрическая модель

Введение случайного компонента в экономическую модель приводит к тому, что взаимосвязь остальных ее переменных перестает быть строго детерминированной и становится стохастической, что и наблюдается в реальной действительности. Это отчасти делает модель доступной для эмпирической проверки на основе статистических данных о конкретном экономическом объекте. Если проверка показала адекватность модели, то иногда удается оценить параметры функционирования конкретного экономического объекта и сформулировать рекомендации для принятия практических решений. Работа с эконометрическими моделями требует использования инструментария оценивания и статистической проверки модели (“наука” моделирования), а также решения проблем выбора типа модели, набора объясняющих переменных и вида связей между ними (“искусство” моделирования).

14.3. Экономические данные: перекрестные данные (cross-section data) и временные ряды (time series)

Статистические данные в эконометрике являются основой для выявления и обоснования эмпирических закономерностей. Без конкретных количественных данных, характеризующих функционирование исследуемого экономического объекта, не всегда возможно определить практическую значимость применяемой экономической модели, даже если целью является выявление преимущественно качественных закономерностей.

Экономические данные обычно делят на два вида: перекрестные данные (cross-section data) и временные ряды (time series). Перекрестные данные - это данные по какому-либо экономическому показателю, полученные для разных однотипных объектов (фирм, регионов). При этом либо все данные относятся к одному и тому же моменту времени, либо их временн я принадлежность несущественна. Временные ряды - это данные, характеризующие один и тот же объект, но в различные моменты времени. К первому типу, например, относятся данные бюджетных обследований населения в определенный момент времени; ко второму - данные о динамике уровня инфляции за определенный период. Данные временных рядов характеризуются определенными зависимостями и закономерностями их последовательных значений, например, могут быть связаны между собой последовательные отклонения от общей тенденции развития; в этих связях экономических показателей могут присутствовать задержки (временн е лаги) и т. д. Это обусловливает необходимость специальных методов их обработки и анализа по сравнению с данными перекрестных выборок.

14.4. Цели и методы сбора статистических данных

Целью сбора экономических данных является получение информационной базы для принятия решений. Естественно, что анализ данных и принятие решений проводится на основе какой-либо интуитивной (неявной) или количественной (явной) экономической модели. Поэтому собирают именно те данные, которые необходимы для соответствующей модели.

Существуют различные методы сбора экономических данных: путем опроса, анкетирования и интервьюирования, получения официальной статистической отчетности и т.д. В большинстве стран существуют статистические органы, занимающиеся сбором, обработкой, распространением и публикацией важнейших данных. Этой деятельностью занимаются также многие специализированные государственные и частные агентства.

14.5. Подготовка статистических данных и использование их в модели

При подготовке статистических данных для работы с экономической моделью возникают две проблемы. Во-первых, могут отсутствовать необходимые для модели данные. Во-вторых (если все данные есть), нужно правильно отобрать их для конкретной модели - так, чтобы они были согласованы и имели общую методическую базу оценки. При отсутствии нужных данных они нередко могут быть рассчитаны по имеющимися. Например, если отсутствуют данные о темпе инфляции (*INF*), но имеются данные о дефляторе валового внутреннего продукта (*DEF*), то инфляция (по ВВП) может быть рассчитана (в %):

$$INF = \left(\frac{DEF}{DEF_{-1}} - 1 \right) \cdot 100 \qquad (2)$$

(индекс -1 означает предшествующий год).

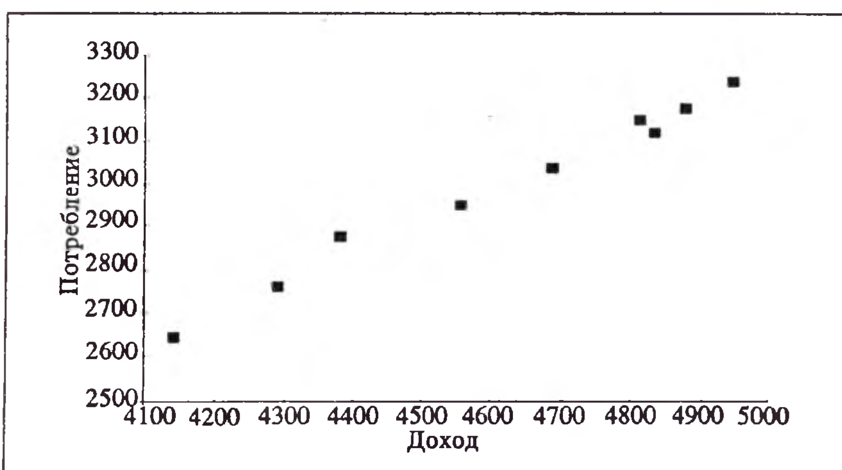
Если все нужные данные есть, то для модели необходимо их преобразовать в некоторый взаимно согласованный набор. Если это данные в денежном выражении, то это должны быть всюду одни и те же текущие, либо фиксированные (одного и того же года), денежные единицы. Реальным объемным показателям (то есть в фиксированных ценах) должны соответствовать реальные относительные показатели (например, процентные ставки нужно скорректировать на темп инфляции). В соответствии с решаемой задачей определяют и обобщающие показатели: валовой национальный продукт (ВНП, или GNP), валовой внутренний продукт (ВВП, или GDP), валовые внутренние или национальные сбережения, дефлятор ВНП или ВВП и т. д.). Например, если речь идет о внутреннем производстве и о влиянии на него внутренних инвестиций, то в качестве обобщающего показателя, на который влияют эти инвестиции, должен выступать ВВП, а не ВНП.

14.6. Различные способы представления экономических данных: таблицы, диаграммы, графики

Собранные данные могут быть представлены в различной форме: в виде таблиц, диаграмм, графиков. Сформулировав в явном виде экономическую модель, например, предположив, что совокупное потребление линейно растет с ростом совокупного дохода, мы должны собрать данные по тем экономическим показателям, которые входят в исследуемую модель, то есть данные по совокупному потреблению и совокупному доходу. Это можно сделать, например, взяв годовые данные из национальных счетов какой-либо страны за некоторый промежуток времени. Эти данные могут быть представлены в виде таблицы: так, в таблице приведены данные по ВВП ("доход") и объему личных потребительских расходов ("потребление") в США, в млрд. долларов в ценах 1987 г.

Год	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Доход	3860	4114	4243	4362	4497	4674	4801	4840	4784	4907
Потребление	2533	2657	2772	2878	2961	3075	3141	3178	3161	3243

Эти данные могут быть представлены также в виде точек на координатной плоскости (диаграммы рассеивания):



Подготовленные данные могут быть подставлены в теоретическую модель, представленную аналитически (в виде некоторой математической модели, например уравнения $CONS = a + b \cdot GDP + \varepsilon$) или в графическом виде (например, прямой линии на плоскости *CONS GDP*). Здесь *CONS* - потребление, *GDP* - ВВП (доход). При этом возникает ряд проблем, важнейшими из которых являются проверка согласованности теоретической модели с данными, оценка параметров модели и проверка предположений (гипотез), лежащих в основе модели.

14.7. Проверка экономических моделей: оценивание коэффициентов, проверка гипотез

Проверить экономическую модель - это значит, в первую очередь, определить, насколько она согласуется с реальными данными об изучаемом объекте. Для этого по эмпирическим данным вычисляются различные статистические характеристики, позволяющие оценить количественно параметры модели, проанализировать надежность этих оценок, проверить различные гипотезы, лежащие в основе исследуемой модели.

14.8. Построение теоретических моделей на основе экономических данных

Располагая экономическими данными, можно не только оценить параметры в уже имеющейся модели, но и выявить эмпирически неизвестные ранее закономерности. А затем уже, на основе выявленных закономерностей построить соответствующую теоретическую модель.

14.9. Статистические методы

Задачей экономического исследования является уяснение природы экономического объекта, раскрытие механизма взаимосвязи между важнейшими его переменными. Такое понимание позволяет разработать и осуществить необходимые меры по управлению данным объектом, или экономическую политику. Для этого нужны адекватные задаче методы, учитывающие природу и специфику экономических данных, служащих основой для качественных и количественных утверждений об изучаемом экономическом объекте или явлении.

14.10. Стохастическая природа экономических данных

Любые экономические данные представляют собой количественные характеристики каких-либо экономических объектов. Они формируются под действием множества факторов, не все из которых доступны внешнему контролю. Неконтролируемые факторы могут принимать случайные значения из некоторого множества значений и тем самым обусловливать случайность данных, которые они определяют. Стохастическая природа экономических данных обусловливает необходимость применения специальных адекватных им статистических методов для их анализа и обработки.

14.11. Понятие случайной переменной

Фундаментальными понятиями статистического анализа являются понятия вероятности и случайной величины (переменной). Случайной переменной мы называем переменную, которая под воздействием случайных факторов может с определенными вероятностями принимать те или иные значения из некоторого множества чисел. Это переменная, которой (даже при фиксированных обстоятельствах) мы не можем приписать определенное значение, но можем приписать несколько значений, которые она принимает с определенными вероятностями. Под вероятностью некоторого события (например, события, состоящего в том, что случайная переменная приняла определенное значение) обычно понимается доля числа исходов, благоприятствующих данному событию, в общем числе возможных равновероятных исходов. Категория "равновероятные исходы" не определяется, а принимается интуитивно. Например, при бросании монеты выпадение орла и решки считается равновероятным (вероятность каждого равна ½), а случайная величина числа "орлов" при одном бросании монеты может быть равна 0 или 1 с вероятностями ½.

Совокупность значений $\{x_k\}$ случайной величины X и вероятностей $\{P_k\}$, с которыми она их принимает, называют *законом распределения случайной величины*. Функция $P(x)$, как и любая функциональная зависимость, может быть представлена в форме таблицы, формулы или графика. Например, закон распределения числа очков при бросании игрального кубика может быть представлен в виде таблицы:

X	1	2	3	4	5	6
P	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Очевидно, что сумма всех этих вероятностей должна равняться единице, поскольку считаем, что с вероятностью "единица" переменная принимает хоть какое-нибудь из этих значений. Обычная (неслучайная, или детерминированная) переменная является предельным случаем случайной переменной, принимая единственное (при фиксированных обстоятельствах) значение с вероятностью "единица".

Различают дискретные и непрерывные случайные величины. Случайная величина *дискретна*, если результаты наблюдений представляют собой конечный или счетный набор возможных чисел. Случайная величина непрерывна, если ее значения могут лежать в некотором континууме возможных значений. (Это предполагает, что их нельзя пересчитать, ставя в соответствие им натуральные числа 1, 2, ...) Значения непрерывной случайной величины могут лежать на отрезке, интервале, луче и т. д.

14.12. Понятия генеральной совокупности и выборки (выборочной совокупности)

В основе математической статистики лежат понятия генеральной совокупности и выборки (выборочной совокупности).

Под *генеральной совокупностью* мы подразумеваем все возможные наблюдения интересующего нас показателя, все исходы случайного испытания или всю совокупность реализаций случайной величины X. Пример генеральной совокупности - данные о доходах всех жителей какой-либо страны, о результатах голосования населения по какому-либо вопросу и т.д. Однако в большинстве случаев мы имеем дело только с частью возможных наблюдений, взятых из генеральной совокупности, и называем это множество (точнее подмножество) значений выборкой. Таким образом, *выборка* - это множество наблюдений, составляющих лишь часть генеральной совокупности. Выборка объема n - это результат наблюдения случайной величины в вероятностном эксперименте, который повторяется n раз в одних и тех же условиях (которые могут контролироваться), а следовательно, и при неизменном распределении случайной величины X. Процесс, который приводит к получению выборочных данных, называют выборочным исследованием.

Мы обычно говорим о генеральной совокупности, когда используем определенные теоретические модели, но на практике в нашем распоряжении имеются лишь выборочные данные, и поэтому мы можем строить оценки теоретических характеристик, основываясь лишь на данных выборочных наблюдений. Мы обсудим соотношение между теоретическими характеристиками и их выборочными оценками позднее. Подчеркнем лишь, что целью математической статистики является получение выводов о параметрах, виде распределения и других свойствах случайных величин (генеральной совокупности) по конечной совокупности наблюдений - выборке.

Выборку называют репрезентативной (представительной), если она достаточно полно представляет изучаемые признаки и параметры генеральной совокупности. Для репрезентативности выборки важно обеспечить случайность отбора, с тем чтобы все объекты генеральной совокупности имели равные вероятности попасть в выборку. Для обеспечения репрезентативности выборки применяют следующие способы отбора: простой отбор (последовательно отбирается первый случайно попавшийся объект), типический отбор (объекты отбираются пропорционально представительству различных типов объектов в генеральной совокупности), случайный отбор - например, с помощью таблицы случайных чисел и т.п.

14.13. Обработка экономических данных

Целью статистического исследования является обнаружение и исследование соотношений между статистическими (экономическими) данными и их использование для прогнозирования и принятия лучших решений. Под термином "статистические данные" мы будем подразумевать набор наблюдаемых значений одной или нескольких переменных, характеризующих изучаемое явление или рассматриваемый экономический объект.

Экономические данные, записанные в порядке их регистрации, обычно труднообозримы и неудобны для дальнейшего анализа. Задачей статистического описания данных является получение такого их представления, которое позволяет наглядно выявить их вероятностные характеристики. Для этого применяются различные формы упорядочивания данных - по возрастанию, по совпадающим значениям, по интервалам и т.п.

14.14. Дискретные случайные величины

Процедуру обработки дискретных выборочных данных рассмотрим на примере объема продаж холодильников в супермаркете за 10 рабочих дней. Пусть объемы продаж указаны в первой строчке нижеприведенной таблицы.

Данные наблюдений представляют собой выборку, состоящую из $n = 10$ наблюдений. Простейшим способом организации данных в выборке является их группировка по возрастанию - данные при этом упорядочиваются по величине, т.е. записываются в виде последовательности $x^{(1)}, x^{(2)},..., x^{(n)}$, в которой $x^{(1)} \leq x^{(2)} \leq ... \leq x^{(n)}$. Последовательность упорядоченных по величине данных приведена во второй строчке таблицы. Разность между максимальным и минимальным элементами выборки $x^{(n)}-x^{(1)}=S$ называется размахом выборки.

$x_1, x_2, \dots, x_n$	1,5,5,6,2,5,6,2,6,5					n=10
$x^{(1)} \le x^{(2)} \le \dots \le x^{(n)}$	1,2,2,5,5,5,5,6,6,6					S=5
$z_1 \ne z_2 \ne \dots \ne z_k$	z_k	1	2	5	6	$\sum_k$
абсолютные частоты	n_k	1	2	4	3	10 = n
относительные частоты	ω_k	0,1	0,2	0,4	0,3	1
накопленные частоты	Ω_k	0,1	0,3	0,7	1	-
функция распределения	$F_k(z)$	0	0,1	0,3	0,7	-

Следующим этапом организации выборки является подсчет частот, с которыми встречаются различные элементы выборки z_1, $z_2, \dots, z_k$, где $k \le n$ - число различных чисел, содержащихся в выборке. Данная выборка содержит 4 различных числа: $z_1 = 1$, $z_2 = 2$, $z_3 = 5$, $z_4 = 6$.

Пусть число z_j встречается в выборке n_j раз, тогда число n_j называется частотой или абсолютной частотой элемента выборки z_j. Эти частоты приведены в четвертой строчке таблицы. Очевидно, что сумма абсолютных частот равна числу наблюдений:

$$\sum_{j=1}^{k} n_j = n.$$

От абсолютных частот удобнее перейти к относительным, определяемым по отношению к объему выборки n:

$$\omega_j = \frac{n_j}{n}.$$

Очевидно, что сумма относительных частот равна единице, т.е.

$$\sum_{j=1}^{k} \omega_j = 1.$$

Последовательность пар (z_j, ω_j) называют статистическим распределением выборки. Обычно статистическое распределение записывается в виде таблицы, первая строка которой содержит различные элементы выборки z_j, а вторая - их относительные частоты ω_j.

При неограниченном росте числа наблюдений относительные частоты значений z_j стремятся к вероятностям $P_j = \text{Prob}\{X=z_j\}$, а статистическое распределение выборки переходит в закон распределения дискретной случайной величины X.

Получение статистического распределения объема продаж важно для определения наиболее вероятных объемов продаж и, следовательно, соответствующих запасов товара.

Наряду с частотами подсчитываются также накопленные частоты:

$$\sum_{j=1}^{m} n_j = N_m,$$

которые показывают, сколько раз в выборке встречаются значения, меньшие или равные данной величине, и накопленные (кумулятивные) относительные частоты:

$$\sum_{j=1}^{m} \omega_j = \Omega_m,$$

приведенные в пятой строчке таблицы.

Вместо кумулятивных частот часто подсчитывают выборочную функцию распределения $F_n(x)$, определяемую по значениям накопленных частот:

$$F_n(x) = \sum_{z_j < x} \omega_j = \frac{1}{n} \sum_{z_j < x} n_j, \tag{3}$$

где суммируются частоты только тех элементов выборки, для которых выполняется неравенство $z_j < x$. Значения выборочной функции распределения приведены в последней строчке таблицы. Ее отличие от кумулятивной частоты состоит в том, что она показывает, какое относительное число раз в выборке встречаются значения, меньшие данной величины (а не меньшие или равные). Выборочная функция распределения представляет собой кусочно-постоянную неубывающую функцию, обращающуюся в нуль при $x \le x^{(1)}$ и принимающую значение "единица" при $x > x^{(n)}$.

14.15. Сравнение относительных частот в выборке и в генеральной совокупности. Репрезентативность выборки

Проводя какой-либо вероятностный эксперимент, например подбрасывая монету N раз и подсчитывая число определенных исходов этого эксперимента, скажем, число выпадений орла $N_{\text{орла}}$, мы можем определить частоту появления данного исхода (“орла”) в серии испытаний как отношения числа испытаний, в которых выпал “орел”, к общему числу испытаний $\left(\frac{N_{\text{орла}}}{N}\right)$. В общем случае мы можем дать следующее определение.

Относительной частотой появления события $\nu(A_k)$ называется отношение числа опытов N_k, в которых произошло событие A_k, к полному числу испытаний N:

$$\nu(A_k) = \frac{N_k(A_k)}{N}. \tag{4}$$

Проводя достаточно большое число опытов, мы можем заметить, что вначале, при малом числе опытов, частота появления какого-либо события, казалось бы, ведет себя случайным образом, но с увеличением числа испытаний ее значение стабилизируется, стремясь к определенному пределу, который и называется вероятностью этого события. Формально, такое, вообще говоря, нестрогое определение вероятности $P(A_k)$ события A_k записывается так:

$$P(A_k) = \lim_{N \to \infty} \frac{N_k(A_k)}{N}, \tag{5}$$

если указанный предел существует.

Такое определение вероятности имеет смысл только при устойчивости частоты. Так, английский статистик Пирсон, подбросив монету 12000 раз, нашел, что частота появления "решки" составила при этом приблизительно 0,5069, а для 24000 бросаний - 0,5005, что приближается к классическому результату 0,5.

Рассмотрим следующий простой пример - бросание игрального кубика. В этом случае вероятности (P) выпадения любого числа очков (X) от 1 до 6 одинаковы и равны $^1/_6$. Пусть генеральной совокупности соответствует распределение в верхней таблице, а некоторая выборка из нее представлена эмпирическим распределениями - в нижней:

X	1	2	3	4	5	6
P	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$

x_k	1	2	3	4	5	6
w_k	0,16	0,17	0,17	0,16	0,17	0,17

Из таблиц видно, что относительные частоты в выборке близки к относительным частотам-вероятностям генеральной совокупности. Требование близости соответствующих частот соответствует понятию репрезентативности выборки.

Аналогично можно рассуждать в рассмотренном выше примере с дневными объемами продаж холодильников. Если рассматривать $z_k=k$ (дневное число продаж) в качестве значений случайной переменной Z, то при достаточно большом числе наблюдений относительные частоты появления значений z_k будут стремиться к вероятности

$$Prob\{Z = z_k\} = \lim_{N \to \infty} \frac{N_k\{Z = z_k\}}{N}, \tag{6}$$

а относительные накопленные частоты - к вероятности

$$Prob\{Z < z\} = \sum_{z_i < z} Prob\{Z = z_i\} = F_Z(z) = \lim_{N \to \infty} \frac{N\{Z < z_k\}}{N}. \tag{7}$$

которая является функцией конкретного значения z и называется функцией распределения дискретной случайной величины Z.

Здесь и далее индекс Z (большая буква латинского алфавита) поясняет, какую случайную величину описывает соответствующая функция (мы будем использовать его в основном в определениях и далее опускать, если это не вызывает недоразумений), а z (малая буква латинского алфавита) - аргумент функции, принимающий значения из множества всевозможных реализаций случайной величины Z.

14.16. Непрерывные случайные величины. Группировка выборочных данных по интервалам значений. Построение гистограммы

При больших объемах выборки, содержащей значения некоторой непрерывной случайной величины, ее элементы группируют по интервалам значений. Для этого интервал выборки, содержащий все ее значения, разбивают на k непересекающихся интервалов, длина которых для удобства расчетов чаще всего выбирается одинаковой и равной размаху выборки, деленному на желаемое число интервалов:

$$\Delta x = \frac{s}{k} \equiv \frac{x^{(n)} - x^{(1)}}{k}. \tag{8}$$

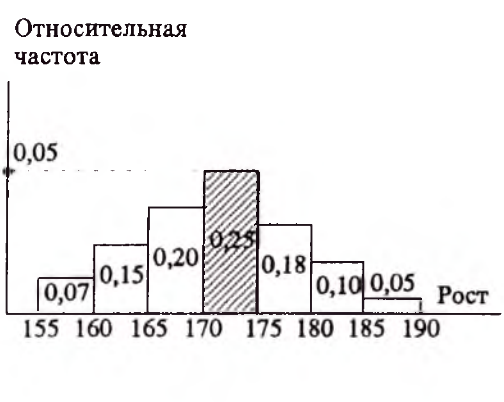
После того, как частичные интервалы выбраны, так же, как и в "точечном" случае, определяют частоты - количество элементов выборки n_j, попавших в j-й интервал, причем элемент, совпадающий с верхней границей интервала относят к последующему интервалу.

Наряду с частотами подсчитываются относительные частоты, накопленные частоты и накопленные относительные частоты. Полученные результаты также записывают в виде таблицы, первая строка которой содержит границы последовательных интервалов, а вторая - соответствующие им частоты (абсолютные, относительные, накопленные и накопленные относительные частоты). Как и для "точечной" выборки, для выборки, сгруппированной по интервалам по значениям накопленных частот, может быть построена выборочная функция распределения. Для наглядного представления выборки часто используют ее графическое отображение - гистограмму частот и гистограмму относительных частот. Любая из этих гистограмм представляет собой кусочно-постоянную функцию, принимающую значения $\frac{n_j}{\Delta x}$ или $\frac{\omega_j}{\Delta x}$ на j-м интервале упорядоченной по возрастанию выборки. Эту функцию представляют в виде ступенчатой фигуры, состоящей из прямоугольников шириной Δx и высотой $\frac{n_j}{\Delta x}$ $\left|\frac{\omega_j}{\Delta x}\right|$, построенных на соответствующих интервалах. Площадь j-го прямоугольника равна $\Delta x \cdot \left[\frac{n_j}{\Delta x}\right]$ (или ω_j), а площадь всей ступенчатой фигуры под графиком гистограммы равна объему выборки (для гистограммы частот) или единице (для гистограммы относительных частот).

В качестве примера рассмотрим гистограмму распределения по росту людей в какой-либо выборке, например студентов одного из институтов:

Рост h	155-16	160-16	165-17	170-17	175-18	180-18	185-19
n_h/n	0,07	0,15	0,20	0,25	0,18	0,10	0,05

Высота каждого столбика в изображенной на рисунке гистограмме пропорциональна количеству людей, рост которых попадает в соответствующий интервал. Допустим, что у 250 из 1000 выбранных для обследования студентов рост находится в пределах от 170 до 175 см ($170 \le h \le 175$). Тогда на гистограмме высота столбика,



S_Σ=0,07+0,15+0,20+0,25+0,18+0,10+0,05=1 соответствующего этому интервалу, равна $\frac{n_j}{n\Delta h} = \frac{\frac{n(170 \le h \le 175)}{n}}{\Delta h}$

$= \frac{250}{1000 \cdot 5} = 0{,}05$, а площадь под этим столбиком - $\frac{n_j}{n} = 0{,}25$.

Аналогично в рассмотренном выше примере, если рассматривать рост студентов h в качестве значений случайной переменной H, то при достаточно большом числе наблюдений относительные частоты появления значений h_k в интервале $h \le H < h+\Delta h$ будут стремиться к вероятности попадания значений роста в вышеуказанный интервал $\mathrm{Prob}\{h \le H < h+\Delta h\}$, а относительные накопленные частоты к вероятности $\mathrm{Prob}\{H < h\}$, которая является функцией конкретного значения h и называется функцией распределения $F_H(h)$ непрерывной случайной величины H.

Если длину интервала Δh устремить к нулю, то вероятность попадания в каждый конкретный интервал также будет стремиться к нулю. Однако отношение этой вероятности к длине интервала стремится при этом к постоянной величине, называемой плотностью вероятности. Плотность вероятности можно интерпретировать как вероятность попадания реализации случайной величины H в бесконечно малый интервал, содержащий точку h, в расчете на единицу его длины:

$$f(h) = \lim_{\Delta h \to 0} \frac{Prob\{h \le H < h + \Delta h\}}{\Delta h}. \tag{9}$$

График функции $f(h)$ будет гладким (для гладких F), в отличие от графика относительной частоты попадания в единичный интервал.

Более строго, если случайная величина является непрерывной, т.е. принимает любые значения из некоторого интервала, то для нее уже нельзя определить вероятность того, что она принимает некоторое конкретное значение (точечную вероятность). Поскольку в любом конечном интервале содержится бесконечное число значений, то вероятность выпадения одного из них всегда равна нулю. Однако функция распределения случайной величины $F_X(x)$, определяемая как вероятность того, что случайная величина принимает значение меньше данного числа x:

$$F_X(x) = \mathrm{Prob}\{X < x\}, \tag{10}$$

сохраняет смысл и для непрерывной случайной величины. В данном случае это некоторая непрерывная неубывающая функция действительного аргумента x.

14.17. Основные характеристики случайных величин ("статистики")

Для любой случайной величины важную роль, помимо функции распределения, играют числовые характеристики ее распределения, важнейшими из которых являются среднее значение (математическое ожидание случайной величины) и дисперсия. Среднее значение является характеристикой положения частотного распределения, а дисперсия - мерой ширины или разброса распределения. Во многих практических случаях информация о случайных переменных, содержащаяся в частотном распределении является избыточной. Например, для принятия решения о покупке акций важно, в первую очередь, знать средний доход на них и риск инвестирования в них денег, характеризуемый степенью разброса среднего дохода (дисперсией), что эквивалентно знанию положения и ширины частотного распределения возможных доходов на акции.

14.17.1. Среднее значение. Математическое ожидание

Пусть имеется, например, выборка объемов продаж холодильников за 10 дней: $\{x_k\} = \{1,5,5,6,2,5,6,2,6,5\}$. Среднее значение объема продаж за один день для этой выборки мы получим, если сложим все выборочные данные и разделим сумму на их число:

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k = \frac{1+5+5+6+2+5+6+2+6+5}{10} = 4{,}3.$$

Если мы посмотрим на сумму в правой части равенства, то заметим, что многие числа в ней повторяются. При этом число повторений, деленное на общее число данных в выборке, является ни чем иным, как частотой появления соответствующих значений в выборке. Следовательно, среднее значение можно определить и так:

$$\bar{x} = \sum_{\{x_k\}} x_k \cdot w_k = 1 \cdot 0{,}1 + 2 \cdot 0{,}2 + 5 \cdot 0{,}4 + 6 \cdot 0{,}3 = 4{,}3, \qquad (11)$$

где суммирование ведется по всем различным значениям случайной величины X, встречающимся в выборке (в данном примере $\{x_k\} = \{1,2,5,6\}$), а в роли весов выступают частоты этих значений (причем сумма весов равна единице).

В пределе достаточно большого числа наблюдений N частоты w_k значений x_k переходят в соответствующие вероятности P_k = $\text{Prob}\{X=x_k\}$, и дискретная случайная величина X может быть представлена в виде таблицы значений $\{x_k\}$, которые может принимать случайная величина, и соответствующих им вероятностей P_k = $\text{Prob}\{X=x_k\}$:

X	x_1	x_2	...	x_n
P	P_1	P_2	...	P_n

Математическое ожидание или среднее (по генеральной совокупности) значение такой случайной величины определяется как взвешенная сумма всех возможных реализаций случайной величины X, в роли весов в которой выступают вероятности этих реализаций, причем сумма весов равна единице:

$$M[X] = x_1 P_1 + x_2 P_2 + \ldots + x_n P_n = \sum_k P_k \cdot x_k. \qquad (12)$$

Это - числовая характеристика (а не функция, на что указывают квадратные скобки) случайной величины X, что означает, что она соответствует всей величине X, а не различным конкретным ее значениям. Другие обозначения среднего значения: $M[X] \equiv \langle X \rangle \equiv m_X \equiv \mu$.

Математическое ожидание непрерывной случайной величины определяется следующим образом:

$$M[X] = \int_{-\infty}^{+\infty} x dF_X(x) = \int_{-\infty}^{+\infty} x f_X(x) dx \qquad (13)$$

Свойства математического ожидания:

$M[c] = c;$

$M[X+b] = M[X] + b;$

$M[aX] = aM[X];$

$M[aX+b] = aM[X] + b,$

- для любых чисел (констант) a, b и c.

Эти свойства вытекают из определения математического ожидания. Если X и Y - случайные величины, то можно определить новые случайные величины $(X+Y)$, $(X-Y)$, $(X \cdot Y)$, $\left(\frac{X}{Y}\right)$, рассчитав вероятности принятия ими конкретных значений $(x+y)$, $(x-y)$, $(x \cdot y)$ и $\left(\frac{x}{y}\right)$.

Для любых случайных величин X и Y

$$M[X+Y] = M[X] + M[Y]. \qquad (14)$$

Математическое ожидание (ожидаемое, или среднее, значение) часто рассчитывается при сравнении затрат и выгод действия со случайным исходом, например ожидаемого выигрыша в лотерее или ожидаемого дохода на акции или другие рисковые ценные бумаги.

Выборка может рассматриваться как дискретная случайная величина, принимающая значения $x_1, x_2, \ldots, x_n$ с вероятностями $^1/n$ (если некоторые значения x_i совпадают, то для расчета выборочного среднего и других характеристик они могут условно рассматриваться как разные, что не меняет результата), то есть $p_k = {}^1/n$ для всех $1 \le k \le n$. Обозначая выборочное среднее как m_n, имеем

$$m_n[x] \equiv \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k. \tag{15}$$

Нижний индекс n показывает объем выборки, для которой вычисляются выборочные характеристики.

Для разных конкретных выборок, соответствующих одной и той же генеральной совокупности, выборочные средние будут, вообще говоря, различны.

14.17.2. Дисперсия

Когда мы имеем дело со случайной величиной, то, как правило, недостаточно определить только ее среднее значение, но и следует ввести меру ее разброса вокруг среднего значения, характеризующую вариативность значений случайной величины. Так, например, для выборки объемов продаж холодильников важно знать не только средний объем продаж, но и то, в каких пределах он может изменяться ото дня ко дню.

Одной из мер вариативности является дисперсия, определяемая как средний квадрат отклонения случайной величины от среднего значения. Для ее определения мы находим отклонения объемов продаж в каждый из дней от среднего значения, возводим их в квадрат и усредняем (складывая и деля на число наблюдений, уменьшенное на единицу):

$$D_n[x] = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \frac{107}{30} \approx 3{,}57.$$

Используя определение частот появления различных значений в выборке, мы, так же как и в случае среднего значения, можем переписать формулу для расчета дисперсии в виде

$$D_n[x] = \frac{n}{n-1}\sum_{(x_k)}(x_k - \bar{x})^2 \cdot \omega_k = \frac{107}{30} \approx 3{,}57.$$

В пределе достаточно большого числа наблюдений n частоты ω_k значений x_k переходят в соответствующие вероятности $P_k = \mathrm{Prob}\{X = x_k\}$, и для анализа отклонений случайной величины от среднего значения полезно ввести в рассмотрение новую случайную величину $Z = (X - \mu)^2$, значения которой представляют квадраты отклонений случайной величины X от среднего значения $\mu \equiv M[X]$. Распределение этой случайной величины также можно представить в виде таблицы:

Z	$(x_1 - \mu)^2$	$(x_2 - \mu)^2$	...	$(x_n - \mu)^2$
P	P_1	P_2	...	P_n

Математическое ожидание такой случайной величины

$$M[Z] = z_1 \cdot P_1 + z_2 \cdot P_2 + \ldots + z_n \cdot P_n = \tag{16}$$
$$= \sum_k z_k P_k \equiv \sum_k (x_k - \mu)^2 P_k = M[(X - M[X])^2].$$

характеризует среднее отклонение (точнее, средний квадрат отклонения) исходной случайной величины X от среднего значения $M[X] \equiv \mu$ и называется дисперсией случайной величины X, обозначаемой $D(X)$ или σ^2.

Итак, общее определение дисперсии как для дискретной, так и для непрерывной случайной величины имеет вид

$$D[X] = M[(X - M[X])^2] \tag{17}$$

Свойства дисперсии:

$D[c] = 0;$

$D[X+b] = D[X];$

$D[aX] = a^2 D[X];$

$D[aX+b] = a^2 D[X]$

- для любых чисел (констант) a, b и c.

Эти свойства могут быть доказаны из определения дисперсии и свойств математического ожидания.

14.17.3. Связь дисперсии с математическим ожиданием

Для практического расчета дисперсии случайных величин часто бывает удобно использовать следующую формулу:

$$D[X] = M[X^2] - (M[X])^2. \tag{18}$$

Эта формула легко выводится из определения и свойств дисперсии. Раскрывая квадрат разности в (18) и пользуясь свойствами математического ожидания, имеем:

$$D[X] = M[(X - M[X])^2] = M[(X^2 - 2M[X]X + (M[X])^2] =$$
$$= M[X^2] - 2M[X]M[X] + (M[X])^2 = M[X^2] - (M[X])^2.$$

Стандартное отклонение случайной величины σ - мера разброса случайной величины вокруг среднего значения, имеющая размерность данной случайной величины. Если случайная величина x измеряется в $, то величина σ измеряет ее разброс вокруг среднего также в $. Стандартное отклонение - это среднее квадратическое разброса случайной величины, или квадратный корень из ее дисперсии:

$$\sigma[X] = \sqrt{D[X]} \tag{19}$$

Коэффициент вариации случайной величины V - мера относительного разброса случайной величины (безразмерная величина).

$$V[X] = \frac{\sqrt{D[X]}}{|M[X]|}. \tag{20}$$

Коэффициент вариации показывает, какую долю среднего значения случайной величины составляет ее средний разброс.

Дисперсия и другие меры разброса часто применяются при анализе риска различных активов в портфеле и портфеля активов в целом в финансовом анализе, а также при анализе риска других действий со случайным исходом.

Пример расчета числовых характеристик дискретной случайной величины: случайная величина X - количество "решек", выпавших при бросании монеты, распределена по закону $\begin{bmatrix} 0 & 1 \\ ½ & ½ \end{bmatrix}$. Для нее $M[X] = 0 \cdot ½ + 1 \cdot ½ = ½$; $D[X] = M[X^2] - M[X]^2 = (0^2 \cdot ½ + 1^2 \cdot ½) - (½)^2 = ¼$; $\sigma = \sqrt{D[X]} = \sqrt{¼} = ½$; $V[X] = ½ : ½ = 1$.

14.18. Общие свойства случайных величин

Независимо от конкретного распределения случайной величины имеют место общие свойства вероятностных распределений. К ним относятся различного рода неравенства, определяющие границы вероятности попадания случайной величины в заданный интервал, а также утверждения, касающиеся свойств достаточно большого числа случайных величин, - так называемые законы больших чисел.

14.18.1. Неравенство Чебышева

Неравенство Чебышева дает оценку вероятности попадания произвольной случайной величины с известными средним значением $M[X]$ и дисперсией σ^2 в заданный интервал вокруг среднего значения. Согласно этому неравенству

$$P(|X - M[X]| \geq \alpha\sigma) \leq \frac{1}{\alpha^2} \quad \forall \alpha > 0 \text{ или}$$
$$P(|X - M[X]| \leq \alpha\sigma) \geq 1 - \frac{1}{\alpha^2} \quad \forall \alpha > 0. \tag{21}$$

То есть вероятность попадания случайной величины вне интервала вокруг ее среднего значения, пропорционального среднеквадратичному (стандартному) отклонению σ, быстро убывает с увеличением коэффициента пропорциональности (α) и, соответственно, длины этого интервала $2\alpha\sigma$.

Таким образом, неравенство Чебышева наглядно демонстрирует значение стандартного отклонения у как характеристики разброса случайной величины вокруг среднего значения.

Например, при $\alpha=2$ $P(|X - M[X]| \geq 2\sigma) \leq 1/4$, а $P(|X - M[X]| \leq 2\sigma) \geq 3/4$ для любого распределения.

Для нормального распределения:

$P(|X - M[X]| \geq \sigma) = 0{,}32 \leq 1$; $P(|X - M[X]| \leq \sigma) = 0{,}68 \geq 0$;
$P(|X - M[X]| \geq 2\sigma) = 0{,}05 \leq 1/4$; $P(|X - M[X]| \leq 2\sigma) = 0{,}95 \geq 3/4$;
$P(|X - M[X]| \geq 3\sigma) = 0{,}003 \leq 1/9$; $P(|X - M[X]| \leq 3\sigma) = 0{,}997 \geq 8/9$.

Пользуясь неравенством Чебышева, можно оценить вероятность тех или иных отклонений от среднего значения, независимо от природы случайной величины.

14.18.2. Законы больших чисел. Теоремы Бернулли, Ляпунова, Чебышева

Основная особенность случайной величины состоит в том, что нельзя предвидеть, какое значение она примет в результате испытания. Однако при достаточно большом числе испытаний обобщающие характеристики выборок случайных величин практически утрачивают случайный характер. То же верно и в отношении суммы достаточно большого числа случайных величин. При увеличении числа слагаемых в сумме противоположные случайные колебания отдельных величин сглаживаются, и закон распределения суммы приближается при определенных условиях к нормальному распределению. Различные утверждения, относящиеся к этим предельным случаям, носят название законов больших чисел. Первым утверждением такого рода была теорема Бернулли, доказанная им еще в 1713 году.

Теорема Бернулли

При достаточно большом числе независимых испытаний n вероятность того, что сколь угодно малым будет отклонение частоты m/n некоторого события A от вероятности наступления этого события P (при условии, что она постоянна в каждом испытании), стремится к единице, т.е. является почти достоверным событием:

$$\lim_{n\to\infty} Prob \left\{ \left| \frac{m}{n} - P \right| < \varepsilon \right\} = 1 \quad \forall\, \varepsilon > 0. \tag{22}$$

Эта теорема имеет важное значение для статистики и эконометрики, обосновывая выбор частоты осуществления некоторого события в качестве оценки вероятности этого события.

Теорема Ляпунова (центральная предельная теорема)

Распределение суммы n произвольно распределенных и взаимно независимых случайных величин при $n \to \infty$ стремится к нормальному распределению, если вклад отдельных слагаемых в сумму равномерно мал.

Именно эта теорема обосновывает ту огромную роль, которую играет в статистике, эконометрике и во многих других областях знания нормальное распределение. Множество факторов, определяющих тот или иной экономический показатель, как правило, достаточно велико, и при выполнении условий теоремы случайное отклонение этого показателя от среднего значения может быть приближенно описано нормальным распределением.

Теорема Чебышева

При достаточно большом числе n попарно независимых случайных величин с ограниченными дисперсиями ($\sigma^2_k < C$, $k = 1,\dots,n$) вероятность того, что сколь угодно мало отклонение среднего арифметического этих величин от среднего арифметического их математических ожиданий, стремится к единице:

$$\lim_{n\to\infty} Prob \left\{ \left| \frac{X_1 + \ldots + X_n}{n} - \frac{M[X_1] + \ldots + M[X_n]}{n} \right| < \varepsilon \right\} \tag{23}$$

$$\forall\, \varepsilon > 0.$$

Согласно этой теореме, среднее арифметическое достаточно большого числа случайных величин утрачивает характер случайной величины и ведет себя почти как постоянная величина. Последняя теорема имеет особенно важное значение для статистики и эконометрики, обосновывая выбор среднего арифметического выборочных величин в качестве оценки математического ожидания (среднего значения) всей совокупности величин.

Вопросы к главе 14

1. Приведите примеры случайных событий в экономике. Можно ли дать им вероятностное описание? Какой вид вероятности Вы при этом используете?
2. Перечислите различные подходы к определению вероятности. Что общего во всех этих подходах? Чем они различаются?
3. Дайте определение случайной величины. Какова связь между случайными величинами и случайными событиями?
4. В чем отличие случайной переменной от неслучайной (детерминированной)? Какие виды случайных переменных Вы знаете? Приведите примеры.
5. Перечислите основные вероятностные характеристики дискретных случайных величин и дайте их определения. Какова формальная (аналитическая) и геометрическая связь между этими характеристиками?
6. Перечислите основные вероятностные характеристики непрерывных случайных величин и дайте их определения. Какова формальная (аналитическая) и геометрическая связь между этими характеристиками?
7. Какая из величин больше: $\mathrm{Prob}(a < X < b)$ или $\mathrm{Prob}(a \le X \le b)$?
8. Как рассчитать вероятность попадания дискретных и непрерывных случайных величин в интервал: $\mathrm{Prob}(a \le X < b)$:
   а) с помощью функции распределения;
   б) с помощью плотности вероятности для непрерывной случайной величины и с помощью функции вероятности для дискретной случайной величины?
9. Как можно охарактеризовать среднее значение случайной величины? Дайте определение математического ожидания.
10. Перечислите основные характеристики разброса случайных величин и дайте их определения. Какова их связь между собой?
11. Каково различие между вычислением математического ожидания для дискретных и непрерывных случайных величин? Что общего в этих определениях?
12. Докажите основные свойства математического ожидания, исходя из его определения.

13. Дайте подробное определение дисперсии для дискретных и непрерывных случайных величин.
14. Докажите основные свойства дисперсии исходя из ее определения.
15. Выведите формулу связи дисперсии с математическими ожиданиями случайной величины и ее квадрата.
16. В чем состоит основная идея законов больших чисел?
17. В каких случаях применимо неравенство Чебышева?
18. Сформулируйте теоремы Бернулли, Ляпунова и Чебышева.

ГЛАВА 15
РАСПРЕДЕЛЕНИЯ И ВЗАИМОСВЯЗИ СЛУЧАЙНЫХ ВЕЛИЧИН В ЭКОНОМИКЕ

15.1. Теоретический и эмпирический подходы к анализу экономических данных: генеральная совокупность и выборка

Выбрав проблему для изучения и сформулировав экономическую модель, которая может помочь в решении этой проблемы, Вы приходите к необходимости проверки совместимости модели с реальными экономическими данными. При этом следует различать два уровня анализа: теоретический и эмпирический. На теоретическом уровне мы предполагаем, что нам известны все возможные реализации интересующих нас экономических показателей - вся генеральная совокупность. Зная или предполагая статистические свойства генеральной совокупности, мы можем теоретически определить значения параметров в модели, и, соответственно, рассчитать по ней нужные экономические показатели. На практике мы не знаем множества возможных исходов, а наблюдаем только случайно выбранные значения интересующих нас показателей. Располагая лишь выборочными значениями, можно оценить, а не определить точно значения параметров модели; эти оценки будут случайными и меняться от выборки к выборке. Поэтому важно не только знать средние оценки параметров, определенные на основе выборочных данных, но и понимать меры их надежности и случайного разброса, обусловленного случайностью процесса формирования выборки.

Если, например, имеются сведения о доходах и расходах всех жителей некоторой страны, составляющих генеральную совокупность, то можно рассчитать различные статистические характеристики для генеральной совокупности: средние доходы и расходы, показатели вариации (дисперсию, стандартное отклонение, коэффициент вариации) доходов и расходов.

Если же собрать сведения о доходах и расходах части жителей, составляющих выборочную совокупность, то найти истинные статистические характеристики или распределения доходов и расходов, вообще говоря, невозможно. Мы сможем лишь оценить эти величины, а также оценить неточность наших оценок, обусловленных случайной природой процесса получения выборочных данных.

Существуют различные способы сбора статистических данных: перепись, статистическая отчетность предприятий и организаций, опросы, анкетирование, таможенный и налоговый контроль и т.д. В каждом методе сбора данных существуют свои недостатки, обусловливающие погрешность в значениях эмпирических показателей. Эти погрешности можно условно разбить на три группы: систематические ошибки, например занижение полученных доходов, случайные ошибки, обусловленные как выборочной природой собираемых данных, так и неточной (по различным причинам) реакцией субъектов на вопросы, а также ошибки округления. Все эти ошибки надо оценивать и учитывать при статистической обработке данных.

Цель любого оценивания - получить как можно более точное значение неизвестной характеристики генеральной совокупности. Информацией, которой мы при этом располагаем, являются данные выборочного наблюдения. В этих условиях единственным способом построения искомой оценки может быть нахождение такой функции выборочных данных, которая с наибольшей точностью аппроксимирует оцениваемую характеристику генеральной совокупности. В зависимости от способа выражения оценки делятся на точечные оценки, выражаемые одним числом, и интервальные оценки, определяющие числовой интервал, внутри которого может находиться оцениваемый параметр генеральной совокупности.

Пытаясь восстановить характеристики генеральной совокупности, мы сталкиваемся с двумя проблемами: неизвестен как вид распределения генеральной совокупности, так и его параметры. В соответствии с этим существуют два класса оценок: оценки вида распределения и оценки его параметров. В качестве оценки вида распределения (учитывая закон больших чисел) можно взять выборочное распределение, подсчитывая частоту попадания выборочных данных в заданные интервалы, а в качестве оценок параметров распределения генеральной совокупности - соответствующие им выборочные значения. Так, в качестве оценки математического ожидания генеральной совокупности можно взять выборочное среднее, а в качестве оценки дисперсии - выборочную дисперсию.

С момента своего возникновения статистика служила интересам власти. Такие данные, например, как численность населения, пригодного к военной службе, являлись основой для принятия военных решений и решений о количестве оружия, которое следует произвести. Данные национальных счетов экономики служат основой для выработки экономической политики правительством. Статистический анализ очень важен и для принятия решений экономическими субъектами. Так, данные по объемам продаж и другим финансовым показателям деятельности фирмы служат основой для принятия стратегических и тактических решений руководством фирм.

15.2. Основные статистические распределения

При обработке выборочных данных, в силу случайной природы процесса получения выборки, важно знать, каким вероятностным законам подчиняются выборочные значения исследуемого экономического показателя. Существует целый ряд распределений вероятности, которые играют роль эталона в статистических выводах. Это прежде всего равномерное распределение, нормальное распределение (распределение Гаусса) и распределение Стьюдента (t-распределение).

15.2.1. Равномерное распределение

Если значения случайной величины из некоторого интервала можно считать равновероятными, то мы приходим к равномерному распределению случайной величины. Равномерное распределение - это такое распределение вероятности, плотность которого постоянна в заданном интервале изменения случайной величины X: $a \le X \le b$. Равномерно распределенная случайная величина обозначается $\boldsymbol{R}(a,b)$. Там, где встречается $\boldsymbol{R}$ без указания параметров, подразумевается стандартное равномерное распределение на интервале $0 \le X \le 1$: $\boldsymbol{R}(0,1)$.

Плотность вероятности равномерного распределения на интервале $[a, b]$ постоянна на этом интервале:

$$f_U(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & x < a,\ x > b \end{cases},$$

а функция распределения:

$$F_U(x) = \begin{bmatrix} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ 1, & x > b \end{bmatrix}.$$

Для равномерного распределения $M[X] = \dfrac{a+b}{2}$, $D[X] = \dfrac{(b-a)^2}{12}$.

Соответствующие этим функциям графики приведены на рисунке 15.1.

На примере равномерного распределения проще всего показать как графически и аналитически рассчитывать вероятность попадания в заданный интервал, т.е. $\text{Prob}\{x_1 \le X < x_2\}$, используя соотношение между плотностью распределения и функцией распределения. Подобно тому, как масса физического тела, равномерно распреде-

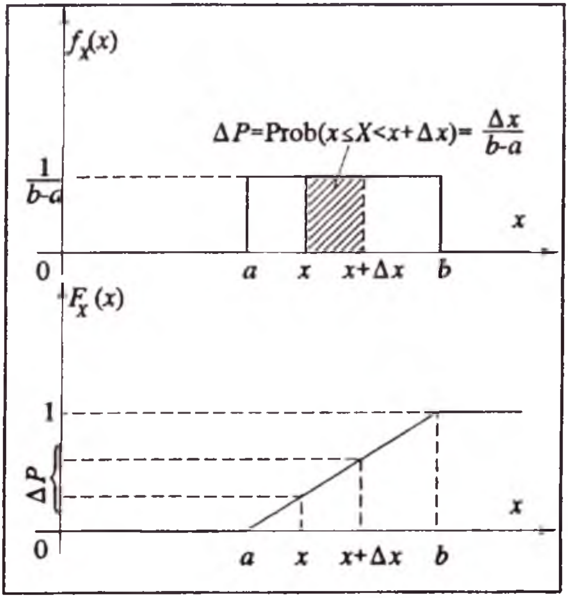


Рисунок 15.1. Плотность вероятности и функция распределения равномерного распределения.

ленная по объему, находится как произведение плотности (массы в единице объема) на объем, так и вероятность попадания равномерно распределенной случайной величины в заданный интервал равна произведению плотности вероятности на длину интервала, и, таким образом, величина вероятности линейно растет с увеличением длины интервала (внутри области определения $[a,b]$).

В общем случае, разбивая интервал значений непрерывной величины $(-\infty, x_2)$ на два интервала $(-\infty, x_1)$ и $[x_1, x_2)$ (одновременные попадания случайной величины в которые являются взаимоисключающими событиями), мы имеем

$$\mathrm{Prob}\{-\infty \le X < x_1\} + \mathrm{Prob}\{x_1 \le X < x_2\} = \mathrm{Prob}\{-\infty \le X < x_2\}.$$

Отсюда находим, что искомая вероятность попадания непрерывной случайной величины в интервал $x_1 \le X < x_2$ равна разности функций распределения этой случайной величины:

$$\mathrm{Prob}\{x_1 \le X < x_2\} = \mathrm{Prob}\{-\infty \le X < x_2\} - \mathrm{Prob}\{-\infty \le X < x_1\} \equiv F_X(x_2) - F_X(x_1).$$

Проводя такие же рассуждения, мы можем найти вероятность попадания непрерывной случайной величины в бесконечно малый интервал $x \leq X < x + dx$:

$$\mathrm{Prob}\{x \leq X < x + dx\} = \mathrm{Prob}\{-\infty \leq X < x+dx\} - \mathrm{Prob}\{-\infty \leq X < x\} \equiv$$
$$\equiv F_X(x + dx) - F_X(x) \equiv dF_X(x) \equiv F'_X(x)dx.$$

В последних двух равенствах мы использовали определение бесконечно малого изменения функции распределения (или дифференциала этой функции). Из найденного соотношения видно, что вероятность попадания непрерывной случайной величины в бесконечно малый интервал $x \leq X < x + dx$ бесконечна мала и пропорциональна величине этого интервала dx. Отношение этой бесконечно малой вероятности к бесконечно малой величине интервала имеет конечное значение и характеризует плотность вероятности в точке x.

Итак, плотность распределения вероятности

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{F_X(x + dx) - F_X(x)}{dx}.$$

И, наоборот,

$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz.$$

На рис. 15.2а приведен характерный график плотности вероятности, а на рис. 15.2б - график соответствующей функции распределения.

Используя выведенную нами взаимосвязь плотности вероятности и функции распределения, несложно показать, что наклон графика функции распределения характеризует плотность вероятности (чем больше плотность вероятности, тем быстрее меняется функция распределения) (точнее $f(x) = \mathrm{tg}(\alpha)$), а площадь под графиком функции плотности вероятности на интервале $x_1 \leq X < x_2$ $\left(\int_a^b f_X(x)dx\right)$ характеризует вероятность попадания непрерывной случайной величины в соответствующий интервал.

При этом суммарная площадь под графиком функции плотности вероятности на всем интервале $-\infty \leq X < +\infty$ равна по определению единице:

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1.$$

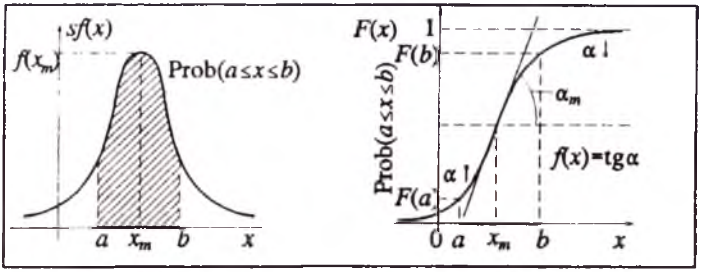


Рис. 15.2а Рис. 15.2б

15.2.2. Нормальное распределение

Если случайная величина формируется под действием большого количества независимых факторов, вклад каждого из которых в значение случайной величины мал, то в силу центральной предельной теоремы эта случайная величина будет иметь нормальное распределение. В роли таких величин могут выступать: объем продаж в конкурентной отрасли или в промышленности в целом, суммарные инвестиции, суммарное потребление домашних хозяйств и тому подобные величины, имеющие аддитивную природу, то есть складывающиеся из многих малых взаимно независимых величин.

Основная особенность случайной величины состоит в том, что нельзя предвидеть, какое значение она примет в результате испытания. Однако при достаточно большом числе испытаний поведение суммы независимых случайных величин почти утрачивает случайный характер и становится почти закономерным. При увеличении числа слагаемых в сумме противоположные случайные колебания отдельных величин сглаживаются и распределение вероятностей суммы становится весьма простым, приближаясь при определенных условиях к нормальному распределению.

Рассмотрим основные свойства нормального распределения. Главное из них - если ряд случайных величин (X_1, X_2, ... , X_n) имеет нормальное распределение, то их сумма ($X_1 + X_2 + ... + X_n$) или любая линейная комбинация ($\alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_n X_n$) также будет иметь нормальное распределение.

Нормальное распределение одной случайной величины X характеризуется лишь двумя параметрами: средним значением, обычно обозначаемым μ, и стандартным отклонением, обычно обозначаемым σ. Это обычно обозначают так: $X = N(\mu,\sigma)$.

Распределение величины $X = \sum_{k=1}^{n} c_k \cdot X_k$, представляющей собой взвешенную сумму n независимых нормально распределенных случайных величин $X_k = N(\mu_k,\sigma_k)$ с параметрами μ_k и σ_k, также будет иметь нормальное распределение с параметрами $\mu = \sum_{k=1}^{n} c_k \cdot \mu_k$ и $\sigma = \sqrt{\sum_{k=1}^{n} c_k^2 \cdot \sigma_k^2}$. В частности, если все $c_k = \frac{1}{n}$, все μ_k и σ_k одинаковы и равны μ_1 и σ_1 соответственно, то $\mu=\mu_1$, а $\sigma = \frac{\sigma_1}{\sqrt{n}}$. Обозначая $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$, имеем, таким образом, $M[\bar{X}] = M[X]$, $\sigma[X] = \frac{\sigma[X]}{\sqrt{n}}$. Отсюда видно, что разброс среднего арифметического независимых нормально распределенных случайных величин стремится к нулю при неограниченном увеличении числа этих величин. Если, например, взята достаточно большая репрезентативная выборка населения, то средний доход в выборке почти наверняка окажется близким к действительному среднему доходу населения.

График плотности вероятности нормального распределения имеет типичный колоколообразный вид и показан на рис. 15.3. Максимум этой функции находится в точке $x = \mu$, а "растянутость" вдоль оси X определяется параметром σ. Чем меньше значение этого параметра, тем более острый и высокий максимум имеет плотность нормального распределения. Аналитически плотность вероятности нормального распределения на интервале $(-\infty, +\infty)$

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{1}$$

а функция распределения

$$F_N(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \tag{2}$$

$$M[X] = \mu,\ D[X] = \sigma^2,\ V[X] = \frac{\sigma}{|\mu|}.$$

Плотность вероятности нормального распределения (1) пропорциональна величине $\exp\left(-\frac{z^2}{2}\right)$, где z - безразмерная величина, определяемая выражением $z = \frac{(x-\mu)}{\sigma}$. Поэтому плотность нормального распределения достаточно быстро (экспоненциально) убывает при удалении x от среднего значения μ. Случайная величина z имеет нулевое математическое ожидание и единичную дисперсию; это вы-

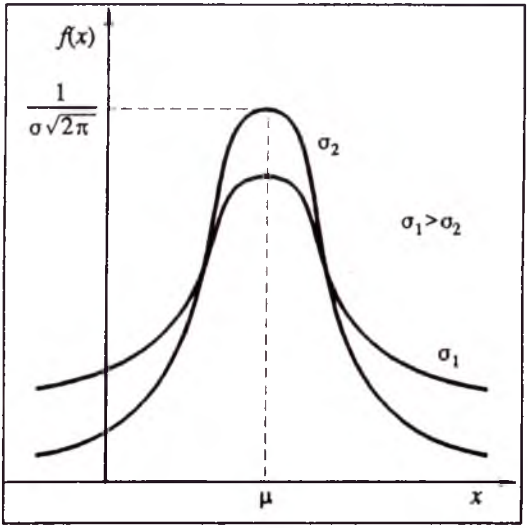


Рисунок 15.3. Функция плотности вероятности нормального распределения

текает из их определений и свойств, учитывая, что $z = \frac{(x - \mu)}{\sigma}$. Она, как и исходная случайная величина x, нормально распределена, но уже не зависит от каких-либо параметров. Поэтому ее распределение может быть протабулировано, то есть значения её плотности вероятности могут быть представлены в виде таблиц. Эта функция называется плотностью стандартного нормального распределения. Стандартное нормальное распределение - это нормальное распределение с параметрами $\mu = 0$ и $\sigma = 1$ ($Z \approx N(0,1)$).

На практике чаще используют таблицы значений не плотности, а функции распределения стандартной нормальной величины $F(z)$. Интересуясь, например, вероятностью того, что нормально распределенная случайная величина X попадает в интервал $x_1 \le X \le x_2$, мы вначале находим соответствующий интервал для нормально распределенной стандартной случайной величины $Z(z_1 \le Z \le z_2)$: $z_1 = \frac{x_1 - \mu}{\sigma}$ и $z_2 = \frac{x_2 - \mu}{\sigma}$. Затем по таблице находим значения функции распределения $F(z_1)$ и $F(z_2)$ и определяем вероятность попадания случайной величины Z в заданный интервал $\text{Prob}\{z_1 \le Z\ z_2\} = F(z_2) - F(z_1)$, совпадающую с искомой вероятностью попадания случайной величины X в заданный интервал $\text{Prob}\{x_1 \le X \le x_2\}$. Геометрически эта вероятность изображается площадью под графиком функции плотности вероятности в интервале от x_1 до x_2.

Аналогично можно решать и обратную задачу - нахождения интервала, в который нормально распределенная случайная величина попадает с заданной вероятностью. Эта процедура часто используется в задачах теории оценивания и проверки гипотез. Так, например, пусть мы хотим проверить гипотезу о равенстве среднего значения нормально распределенной случайной величины м (для генеральной совокупности) нулю, допуская вероятность ошибки 0,05 в случае, если эта гипотеза верна. В этом случае выборочное значение стандартной нормально распределенной случайной величины Z должно попадать в такой интервал, что вероятность $\text{Prob}\{z_1 \le Z \le z_2\} = 0{,}95$. Из этого условия и соображений симметрии можно найти границы интервала - критические значения $z_{кр} = z_2 = -z_1$, такие, что вероятность $\text{Prob}\{z_2 \le Z\} = \text{Prob}\{Z \le z_1\} = \frac{0{,}05}{2} = 0{,}025$. Сравнивая выборочное значение величины $z = \frac{(x - \mu)}{\sigma}$, называемое z-статистикой с критическим значением $z_{кр}$, мы принимаем (если $z_1 \le z \le z_2$) или отвергаем (если $z \le z_1$ или $z_2 \le z$) проверяемую гипотезу с точностью (уровнем значимости) $\varepsilon = 0{,}05$ (5%).

Естественно, что вышеописанную процедуру можно применять, только если известно стандартное отклонение σ или дисперсия σ^2 исследуемой случайной величины, что редко имеет место на практике. Поэтому при оценивании параметров и проверке гипотез чаще применяют другое распределение, являющееся по сути выборочным аналогом нормального распределения и переходящее в него при бесконечно большом числе наблюдений. Это распределение называют распределением Стьюдента или t-распределением.

15.2.3. Распределение Стьюдента

Рассмотрим основные свойства распределения Стьюдента. Во-первых, аналогом безразмерной величины z-статистики, определяемой выражением $z = \frac{(x - \mu)}{\sigma}$, служит также безразмерная величина t-статистика, определяемая выражением $t = \frac{x - \mu}{s}$. В этом выражении вместо стандартного отклонения для генеральной совокупности у стоит выборочное стандартное отклонение s, являющееся, по сути, случайной величиной (меняющейся от выборки к выборке) и определяемое по данным наблюдений x_k с помощью выражения:

$$s_н = \sqrt{\frac{1}{n - 1}\sum_{k=1}^{n} (x_k - \bar{x})^2}.$$

Здесь выборочное среднее обозначено $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$, а через n обозначено число наблюдений.

Во-вторых, в отличие от стандартного нормального распределения, являющегося функцией лишь одной переменной z, t-распределение является не только функцией переменной t, но также зависит от еще одного параметра - числа степеней свободы ν. Число степеней свободы равно общему числу наблюдений, уменьшенному на число линейных связей между ними. Если n выборочных наблюдений связаны s линейными уравнениями, то их распределение имеет $\nu = n-s$ степеней свободы. Линейной связью является, например, формула расчета выборочного среднего $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ и если выборочное среднее входит в формулу какой-либо статистики, то это уменьшает число степеней свободы на единицу.

Распределение Стьюдента имеет случайная величина, равная отношению двух независимых случайных величин: стандартной нормально распределенной величины Z (с нулевым средним значением и единичной дисперсией) и величины $\sqrt{\frac{\chi^2(n)}{n}}$, выражающейся через случайную величину, имеющую распределение χ^2 с n степенями свободы. Распределение χ^2 (хи - квадрат, или распределение Пирсона), имеет сумма квадратов n независимых стандартных нормально распределенных случайных величин (с нулевыми средними значениями и единичными дисперсиями).

Вводя новую случайную величину

$$T(n) = \frac{Z\sqrt{n}}{\sqrt{\chi^2(n)}} = \sqrt{\frac{\frac{\chi^2(1)}{1}}{\frac{\chi^2(n)}{n}}},$$

мы получим для нее t-распределение Стьюдента с n степенями свободы с плотностью вероятности $f(x,n) = B\left(\frac{1+x^2}{n}\right)^{-\frac{n+1}{2}}$. График функции плотности вероятности распределения Стьюдента (рис. 15.4), как и стандартного нормального распределения, имеет симметричный колоколообразный вид, но является более "сплюснутым" по вертикали.

Из симметричности распределения Стьюдента вытекает важное соотношение между критическими точками этого распределения: $t_{\alpha}(n) = t_{1-\alpha}(n)$.

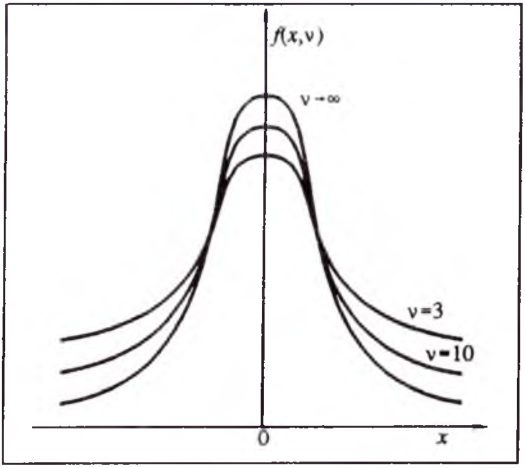


Рисунок 15.4. Плотность вероятности распределения Стьюдента ($M[X] = 0$, $D[X] = \frac{\nu}{\nu - 2}$).

На практике обычно используют не таблицы функции распределения Стьюдента $F(z)$, а таблицы критических точек функции распределения Стьюдента, то есть точек с заданной вероятностью попадания в начинающиеся от них "хвосты" распределения.

Распределение Стьюдента используется, например, при проверке гипотез:

- о среднем значении нормальной генеральной совокупности при неизвестной дисперсии;
- о линейной независимости двух случайных величин (равенстве нулю коэффициента корреляции) - см. ниже в этой главе;
- о статистической значимости коэффициента линейной регрессии.

15.3. Таблицы распределений и их использование. Примеры расчетов вероятности попадания в заданный интервал с помощью таблиц

15.3.1. Работа с таблицами стандартного нормального распределения

Для практического применения приведенных выше распределений к проведению статистических расчетов служат таблицы распределений. Рассмотрим использование таблиц распределений на примере нормального распределения.

Таблица функции распределения стандартного нормального распределения на интервале (-∞, +∞) $F_N(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$ имеет вид

z	0	1	2	3	4	5	6	7	8	9
0,0	0,500	0,503	0,507	0,511	0,515	0,519	0,523	0,527	0,531	0,535
0,1	0,539	0,543	0,547	0,551	0,555	0,559	0,563	0,567	0,571	0,575
...	...	...	...	...	...	...	...	...	...	...
4,0	0,999	0,999	0,999	0,999	-	-	-	-	-	-

В приведенном фрагменте таблицы значения функции распределения приведены с точностью до пятого десятичного знака, а значения аргумента z - до второго десятичного знака. В самом левом столбце таблицы приведены значения z от 0,0 до 4,0 с точностью до десятых долей, а в верхней строке таблицы приведены сотые доли z. Значение функции распределения, соответствующее определенному значению аргумента z (например, 0,14), находится на пересечении строки с десятыми долями z (в данном примере - 0,1) и столбца с сотыми долями z (в данном примере - 4). В рассматриваемом примере оно равно 0,55567 и означает вероятность попадания случайной величины z в полубесконечный интервал (-∞, 0,14). Эта вероятность равна 0,5 при $z = 0$, так как это значение делит область изменения случайной величины z на две равновероятные части, и стремится к единице при увеличении z. Для того, чтобы рассчитать вероятность попадания величины z в конечный интервал $\alpha \le z < \beta$, следует воспользоваться формулой: $\text{Prob}\{\alpha \le Z < \beta\} = F(\beta) - F(\alpha)$. Пусть, к примеру, $\alpha = 0,1$, а $\beta = 0,14$. Тогда по таблице находим $F(0,1) = 0,53983$, а $F(0,14) = 0,55567$. Следовательно, искомая вероятность попадания величины z в интервал [0,1, 0,14) равна $F(0,14) - F(0,1) = 0,55567 - 0,53983 = 0,01584$.

Вспоминая, что стандартная нормальная величина z связана с исходной случайной величиной x соотношением $z = \frac{(x - \mu)}{\sigma}$, мы можем определить вероятность попадания произвольной нормально распределенной величины x в интервал $a \le x < b$ как вероятность попадания стандартной нормальной величины z в интервал $\frac{a - \mu}{\sigma} \le z < \frac{b - \mu}{\sigma}$. Определяя последний интервал, мы можем рассчитать искомую вероятность с помощью таблиц так же, как в рассмотренном выше примере.

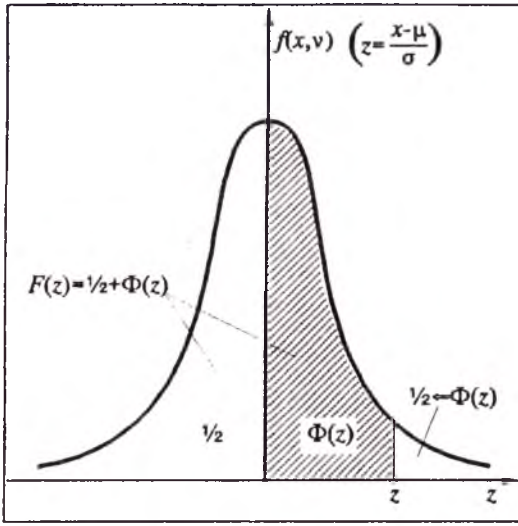


Рисунок 15.5. Функции $f(z)$, $F(z)$, $\Phi(z)$ стандартного нормального распределения

Отметим, что иногда в таблицах стандартного нормального распределения приведена не функция распределения, а величина 1 - $F_N(z) = \frac{1}{\sqrt{2\pi}}\int_z^{+\infty} e^{-\frac{t^2}{2}}dt$, определяющая вероятность попадания случайной величины Z в правый "хвост" распределения (интервал $[z, +\infty)$) и изменяющаяся от 0,5 до 0 при изменении z от 0 до $+\infty$ или величина $\Phi_N(z) = F_N(z) - ½ = \frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{t^2}{2}}dt$, определяющая вероятность попадания случайной величины Z в среднюю часть функции распределения (интервал $[0, z)$) и изменяющаяся от 0 до 0,5 при изменении z от 0 до $+\infty$.

Связь между собой трех вышеупомянутых функций показана на рисунке 15.5.

Для всех этих функций вероятность попадания случайной величины Z в заданный интервал рассчитывается как разность значений соответствующих функций на концах этого интервала.

15.3.2. Работа с таблицами *t*-распределения Стьюдента

В таблице функции распределения Стьюдента приводятся обычно, для различных чисел степеней свободы ν, критические точки, соответствующие приведенным в верхней строке таблицы вероятностям α попадания в правый "хвост" распределения. Иными словами, в приведенной ниже таблицы число α - это вероятность превышения t-статистикой приведенного в таблице критического значения при соответствующем числе степеней свободы v.

Таблица функции распределения Стьюдента имеет вид:

$v \backslash \alpha$	0,005	0,01	0,025	0,05	0,1
1	63,657	31,821	12,706	6,314	3,078
...	...	...	...	...	...
10	3,169	2,764	2,228	1,812	1,372
...	...	...	...	...	...
30	2,750	2,457	2,042	1,697	1,310
∞	2,576	2,326	1,960	1,645	1,282

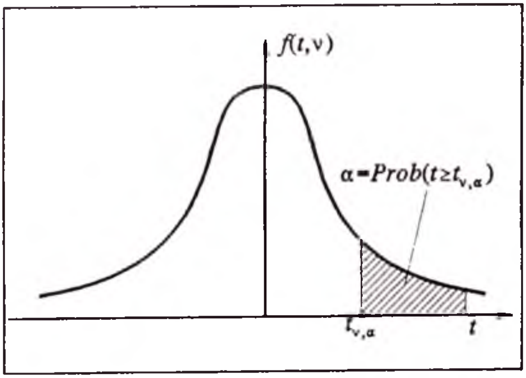


Рисунок 15.6. Односторонняя критическая область распределения Стьюдента

Критическая точка $t_{v,\alpha}$ (например, $t_{10;0,05}$) находится на пересечении строки с числом степеней свободы (в данном случае $v = 10$) и столбца с заданной вероятностью (в данном случае $\alpha = 0,05$). Из приведенной таблицы находим, что $t_{10;0,05} = 1,812$. Напомним, что критическая точка в данном случае имеет следующий смысл:

$$\mathrm{Prob}\{t > t_{v,\alpha}\} = \alpha.$$

Отметим, что иногда таблицы распределения Стьюдента приводятся для двусторонних критических точек $t^{*}_{v,\alpha}$, определяемых из условия $\mathrm{Prob}\{|t| > t^{*}_{v,\alpha}\} = \alpha$.

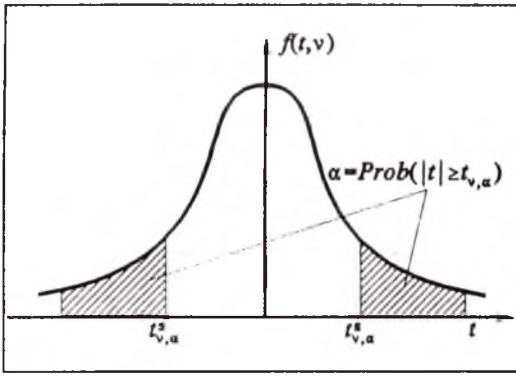


Рисунок 15.7. Двусторонняя критическая область распределения Стьюдента

В силу симметричности распределения Стьюдента эти точки связаны с односторонними критическими точками соотношением $t^{s}_{\nu,\alpha} = t_{\nu,\alpha/2}$, так как при заданной вероятности α попадания в оба “хвоста” распределения вероятность попадания в один из “хвостов” распределения будет в два раза меньше и равна α/2.

Кроме того, в некоторых таблицах распределения Стьюдента вместо малых чисел α (вероятностей попадания в “хвост” распределения) приводятся числа 1-α (вероятности попадания в интервал $(-\infty, t_{\nu,\alpha})$ для односторонних критических точек и в интервал $[-t^{s}_{\nu,\alpha}, t^{s}_{\nu,\alpha})$ для двусторонних критических точек).

15.4. Соотношения между экономическими переменными. Линейная связь. Корреляция

Различные экономические показатели как на микро-, так и на макроуровне не являются независимыми, а связаны между собой; например, цена какого-либо товара и величина спроса на этот товар, объем производства и прибыль фирмы, располагаемый доход и объем личного потребления, инфляция и безработица.

Если не принимать во внимание стохастическую природу экономических данных, то для описания взаимосвязей различных экономических и финансовых показателей между собой применяется функциональный подход. Связь одного из показателей с другими показателями описывается с помощью функций одной $y=f(x)$ или нескольких переменных $y=f(x_1,x_2,\dots,x_n)$. Такой подход применяется там, где вероятностный характер экономических процессов малосущественен для принятия решений.

На самом деле взаимосвязи показателей в экономике редко имеют простой функциональный вид, поскольку на интересующий нас показатель кроме явно учитываемых объясняющих переменных влияет еще множество других факторов, существующих в действительности, но не учитываемых явно в модели; часть из этих факторов - случайные. Это обусловливает стохастическую природу как некоторых экономических переменных, так и взаимосвязей между ними. Стохастические взаимосвязи переменных можно описать с помощью частотных (вероятностных) или корреляционных характеристик.

15.4.1. Вероятностные соотношения: совместная частота (вероятность), условная частота (вероятность), статистическая независимость случайных переменных

Под совместной частотой $\nu(x,y)$ двух случайных величин X и Y мы понимаем относительную частоту события, состоящего в том, что величины X и Y принимают одновременно значения x и y соответственно. В пределе, когда число наблюдений стремится к бесконечности, совместная частота переходит в совместную вероятность $P(x,y)$. Однако совместная частота или вероятность не являются характеристиками именно взаимосвязи случайных переменных, поскольку их значения определяются не только синхронностью изменения исследуемых случайных величин, но и частотой, с которой переменные принимают фиксированные значения. Поэтому для характеристики взаимосвязи случайных переменных чаще используют другую характеристику - условную частоту или вероятность.

Под условной частотой $\nu(y|x)$ двух случайных величин Y и X мы понимаем относительную частоту события, состоящего в том, что величина Y принимает значение y при условии, что величина X уже приняла значение x. В пределе, когда величины X и Y принимают все возможные значения из генеральной совокупности, условная частота переходит в условную вероятность $P(y|x)$.

Если величины X и Y связаны функциональной зависимостью, например, $Y=X^2$, то условная частота (вероятность) $P(y=x^2|x)$ равна единице, так как в 100% случаях (с вероятностью единица) значению X будет соответствовать значение $Y=X^2$.

В противоположном случае, когда величины X и Y совершенно не связаны между собой, условная частота $\nu(y|x)$ никак не будет зависеть от значения x и будет совпадать с частотой появления значения y, $\nu(y)$. При этом совместная частота будет пропорциональна частотам появления значений x и y: $\nu(x,y)=\nu(x)\cdot\nu(y)$. То же относится и к совместной вероятности независимых случайных величин $P(x,y)=P(x)\cdot P(y)$. Данное равенство является формальным определением независимости случайных величин x и y: они называются независимыми, если равенство $P(x,y) = P(x) \cdot P(y)$ выполняется для любых значений x и y.

Если X и Y независимы, то $M[X \cdot Y] = M[X] \cdot M[Y]$, а $D[X+Y] = D[X] + D[Y]$.

15.4.2. Анализ линейной статистической связи экономических данных. Корреляция

В экономических исследованиях одной из основных задач является анализ зависимостей между переменными. Зависимость может быть строгой (функциональной) либо статистической. Алгебра и математический анализ занимаются изучением функциональных зависимостей, то есть зависимостей, заданных в виде точных формул. Но любая такая зависимость в определенной степени является абстракцией, поскольку в окружающем мире, частью которого является экономика, значение конкретной величины не определяется неизменной формулой ее зависимости от некоторого набора других величин. Всегда есть несколько величин, которые определяют главные тенденции изменения рассматриваемой величины, и в экономической теории и практике ограничиваются тем или иным кругом таких величин (объясняющих переменных). Однако всегда существует и воздействие большого числа других, менее важных или трудно идентифицируемых факторов, приводящее к отклонению значений объясняемой (зависимой) переменной от конкретной формулы ее связи с объясняющими переменными, сколь бы точной эта формула ни была. Нахождение, оценка и анализ таких связей, идентификация объясняющих переменных, построение формул зависимости и оценка их параметров являются не только одним из важнейших разделов математической статистики. Это своего рода искусство, учитывающее в каждой конкретной области знаний (в частности, в экономике, о которой идет речь), ее внутренние законы и потребности. Но это также и наука, поскольку выбираемый и оцениваемый вид формулы должен быть объяснен в терминах данной области знаний.

Пусть требуется оценить связь между переменными X и Y (например, связь показателей безработицы и инфляции в данной стране за определенный период времени). В частности, может стоять вопрос, связаны ли между собой эти показатели, и при положительном ответе на него, естественно, встает задача нахождения формулы этой связи. Основой для ответа на этот вопрос являются статистические данные о динамике этих показателей (годовые, квартальные, месячные и т.п.). Эти данные представляют собой некоторую, предположительно - случайную, выборку из генеральной совокупности, то есть из совокупности всех возможных сочетаний показателей инфляции и безработицы в сложившихся условиях.

Таким образом, вывод о наличии связи для всей генеральной совокупности нужно делать по выборочным данным, что само по себе уже делает ответ на поставленный вопрос небезусловным. Более того, по данным выборки ответить на вопрос в приведенной постановке, то есть о наличии связи "вообще", невозможно. Действительно, через любые N точек на плоскости всегда можно провести полином степени N-1 и объявить, что найдена точная формула связи. Однако опыт подсказывает, что если бы мы получили еще одну точку-наблюдение, то она наверняка не удовлетворяла бы найденной формуле. Поэтому вопрос о наличии связи между переменными (в частности - экономическими) следует ставить как вопрос о наличии конкретной формулы (спецификации) такой связи, устойчивой к изменению числа наблюдений. При этом нужно понимать, что ответ на этот вопрос по данным выборки не может быть однозначным и категоричным.

Простейшей формой зависимости между переменными является линейная зависимость, и проверка наличия такой зависимости, оценивание ее индикаторов и параметров является одним из важнейших направлений приложения математической статистики.

Рассмотрим вначале вопрос о линейной связи двух переменных

1) Связаны ли между собой линейно переменные X и Y?

2) Какова формула связи переменных X и Y?

В первом случае переменные X и Y выступают как равноправные, здесь нет независимой и зависимой переменных. Во втором случае речь может идти о нахождении зависимости одной переменной от другой, например об оценивании формулы $Y=a+bX$ (где a и b - неизвестные коэффициенты такой зависимости). В этом случае переменная X является независимой (объясняющей), а переменная Y - зависимой (объясняемой). Вопрос о нахождении формулы зависимости можно ставить после положительного ответа на вопрос о существовании такой зависимости, но эти два вопроса можно решать и одновременно.

Для ответа на поставленные вопросы существуют специальные статистические методы и, соответственно, показатели, значения которых определенным образом (и с определенной вероятностью) свидетельствуют о наличии или отсутствии линейной связи между переменными. В первом случае это коэффициент корреляции величин X и Y, во втором случае - коэффициенты линейной регрессии a и b, их стандартные ошибки и t-статистики, по значениям которых проверяется гипотеза об отсутствии связи величин X и Y.

Вначале объясним логику появления такого показателя, как коэффициент корреляции. Предположим, что между переменными X и Y существует линейная связь. Наличие такой связи можно интерпретировать следующим образом. Если переменная X принимает значения большие, чем ее среднее значение, и связь положительна (на языке формул это означает, что коэффициент b положителен), то значение переменной Y также должно быть больше ее среднего значения и соотношение отклонений X и Y от их средних значений должно быть постоянным. Если в этом случае переменная X принимает значение меньше, чем ее среднее значение, то и значение Y должно быть меньше ее среднего с тем же коэффициентом пропорциональности этих отклонений. Если связь переменных X и Y отрицательна, то положительное отклонение X от среднего значения должно сочетаться с отрицательным отклонением Y от ее средней, а отрицательное отклонение X от среднего значения - с положительным отклонением Y от ее средней - при постоянном соотношении этих отклонений. Если линейной связи между переменными X и Y нет, то положительные отклонения переменной X от ее среднего значения могут (хотя и не обязательно будут) сочетаться как с положительными, так и с отрицательными отклонениями Y от ее среднего, то же можно сказать и про отрицательные отклонения X от среднего.

15.4.3. Коэффициент корреляции для выборки и генеральной совокупности

В качестве меры для степени линейной связи двух переменных используется коэффициент их корреляции. Приведем вначале формулу выборочного коэффициента корреляции переменных X и Y:

$$r_n^*[x,y] = \frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2}\sqrt{\frac{1}{n}\sum_{m=1}^{n}(y_m - \bar{y})^2}}. \qquad (3)$$

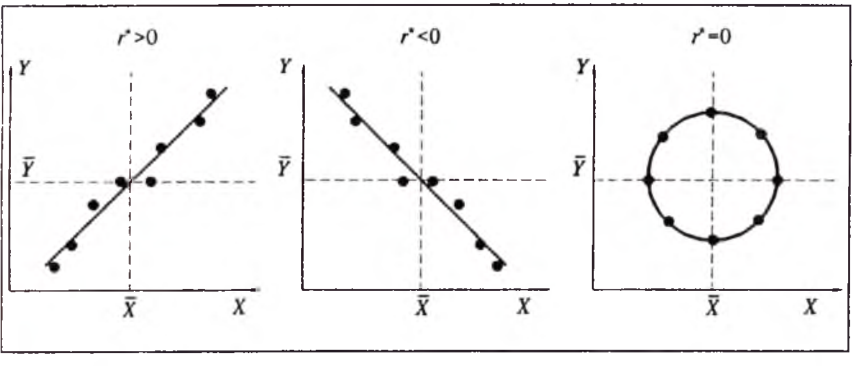


Рисунок 15.8. Типы зависимостей и коэффициент корреляции

По формуле коэффициента корреляции видно, что он будет положителен, если отклонения переменных X и Y от своих средних значений имеют, как правило, одинаковый знак, и отрицательным - если разные знаки.

Коэффициент корреляции является безразмерной величиной (так как размерности числителя и знаменателя есть размерности произведения X Y); его величина не зависит от выбора единиц измерения обеих переменных. Величина коэффициента корреляции меняется от -1 в случае строгой линейной отрицательной связи до +1 в случае строгой линейной положительной связи. Случаи положительной и отрицательной корреляции переменных (с близкими по модулю к единице коэффициентами корреляции) показаны на рис. 15.8. Близкая к нулю величина коэффициента корреляции говорит об отсутствии линейной связи переменных, но не об отсутствии связи между ними вообще. Это ясно из правой части рис. 15.8, где X и Y, очевидно, связаны друг с другом (лежат на одной окружности), но их коэффициент корреляции близок к нулю. Последнее вытекает их того, что каждой паре одинаковых отклонений переменной X от ее среднего значения соответствуют равные по абсолютной величине положительное и отрицательное отклонения переменной Y от ее среднего. Соответственно, произведения этих отклонений "гасят" друг друга в числителе формулы коэффициента корреляции, и он оказывается близким к нулю. Заметим, что в числителе формулы для выборочного коэффициента корреляции величин X и Y стоит их *показатель ковариации*:

$$\mathrm{cov}_n^*[x,y] = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y}). \qquad (4)$$

Этот показатель, как и коэффициент корреляции, характеризует степень линейной связи величин X и Y, и он также равен нулю, если эти величины независимы. Однако, в отличие от коэффициента корреляции, показатель ковариации не нормирован - он имеет размерность, и его величина зависит от единиц измерения величин X и Y. В статистическом анализе показатель ковариации сам по себе используется редко; он фигурирует обычно как промежуточный элемент расчета коэффициента корреляции.

Мы вели до сих пор речь о выборочном коэффициенте корреляции величин X и Y, который рассчитывается для оценки степени линейной связи этих величин по данным выборки. При этом истинным показателем степени линейной связи величин X и Y для закона распределения, имеющегося на генеральной совокупности, является теоретический коэффициент корреляции ρ_{XY}, оценкой которого является выборочный коэффициент корреляции. Коэффициент корреляции для генеральной совокупности определяется следующим образом:

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}. \qquad (5)$$

Стоящий в числителе этой формулы показатель ковариации величин X и Y определяется следующим образом:

$$\mathrm{cov}[X,Y] = M[(X-M[X])\cdot(Y-M[Y])]. \qquad (6)$$

Используя показатель ковариации, удобно записать формулу для дисперсии суммы случайных величин X и Y:

$$D[X+Y] = D[X] + D[Y] + 2\cdot\mathrm{cov}[X,Y]. \qquad (7)$$

Исходя из определения коэффициента корреляции, покажем, что он равен 1 или -1 при строгой линейной зависимости величин X и Y и равен нулю в случае их независимости.

Пусть $Y=a+bX$. Тогда

$$\begin{aligned} M[(X - M[X])\cdot(Y - M[Y])] &= M[(X - M[X])\cdot(a + bX - M[a + bX])] = \\ &= M[(X - M[X])\cdot(a + bX - M[a] - M[bX])] = \\ &= M[(X - M[X])\cdot(a + bX - M[a] - M[bX])] = \\ &= M[(X - M[X])\cdot b(X - M[X])] = bD[X]. \end{aligned}$$

Очевидно также, что $D[Y]=D[a+bX]=b^2D[X]$, и $\rho_{XY} = \frac{\mathrm{cov}[X,Y]}{\sigma[X]\cdot\sigma[Y]} = \frac{b}{|b|}$, то есть коэффициент корреляции равен 1 при положительном коэффициенте b и равен -1 при отрицательном b. Если X и Y независимы, то

$$\mathrm{cov}[X,Y] = M[(X - M[X])\cdot(Y - M[Y])] = M[X - M[X]]\cdot M[Y - M[Y]] = 0\cdot 0 = 0,$$

но необязательно наоборот.

Итак, равный нулю коэффициент корреляции для генеральной совокупности говорит об отсутствии линейной связи рассматриваемых величин. Однако он не свидетельствует об отсутствии их связи вообще. В случае равенства нулю показателя корреляции, например, величин уровней инфляции и безработицы (а это действительно практически так для периода 1970-х - 1980-х годов для экономики США) нужно не говорить сразу о независимости этих показателей в данный период, а попытаться построить более сложную модель их связи, учитывающую, возможно, как нелинейность самой зависимости, так и наличие в ней запаздываний во времени (лагов), а также инерционность динамики соответствующих величин.

15.4.4. Оценивание параметров и проверка гипотез о корреляции случайных переменных

Далее, в анализе коэффициента корреляции возникает следующий вопрос. Если он равен нулю для генеральной совокупности, это вовсе не значит, что он в точности будет равен нулю для выборки. Наоборот, он обязательно будет отклоняться от истинного значения, но чем больше такое отклонение, тем менее оно вероятно при данном объеме выборки. Таким образом, при каждом конкретном значении коэффициента корреляции величин X и Y для генеральной совокупности выборочный коэффициент корреляции является случайной величиной. Следовательно, случайной величиной является также любая его функция, и требуется указать такую функцию, которая имела бы одно из известных распределений, удобное для табличного анализа. Для выборочного коэффициента корреляции r такой функцией является t-статистика, рассчитываемая по формуле $t = r\cdot\frac{\sqrt{n-2}}{\sqrt{1-r^2}}$ и имеющая распределение Стьюдента с (n-2) степенями свободы. Число степеней свободы меньше числа наблюдений на 2, поскольку в формулу выборочного коэффициента корреляции входят средние выборочные значения X и Y, для расчета которых используются две линейные формулы их зависимости от наблюдений случайных величин. Сразу уточним, что для коэффициента корреляции будет проверяться нулевая гипотеза, то есть гипотеза о равенстве его нулю в генеральной совокупности. Эта гипотеза отвергается, если выборочный коэффициент корреляции слишком далеко отклонился от нулевого значения, то есть произошло событие, которое было бы маловероятным в случае $\rho_{XY}=0$.

Здесь, конечно, очень важно понять, что конкретно значат слова “слишком далеко” и “маловероятное событие”. В последнем случае нужно задать вероятность такого события, которая называется в статистике “уровень значимости”. Чаще всего задается уровень значимости 1% или 5%. Если для некоторого показателя проверяется гипотеза о том, что его истинное значение равно нулю, то данная гипотеза отвергается в том случае, если оценка показателя по данным выборки такова, что вероятность получения такого или большего (по модулю) ее значения меньше, чем 1% или 5% соответственно.

На рис. 15.9 дана иллюстрация проверки нулевой гипотезы для коэффициента корреляции, которая может быть использована для рассмотрения общей схемы проверки статистических гипотез. Здесь H_0 - гипотеза о том, что истинное значение коэффициента корреляции равно нулю, альтернативная ей гипотеза H_1 - что оно не равно

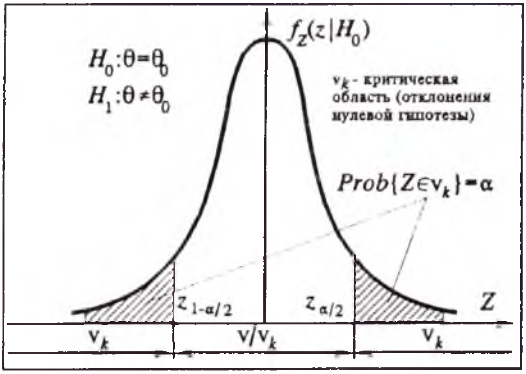


Рисунок 15.9. Проверка нулевой гипотезы для коэффициента корреляции

нулю. Функция f_Z - функция плотности вероятности распределения Стьюдента в случае, если нулевая гипотеза верна (она максимальна при Z=0, где Z - случайная величина выборочного коэффициента корреляции). Заштрихованная область - это область больших по абсолютной величине (маловероятных при выполнении гипотезы H_0) значений выборочного коэффициента корреляции. Если последнее все-таки попало в эту область, то H_0 отвергается. Площадь заштрихованной области, равная α, - уровень значимости, или вероятность того, что туда попадет величина Z при выполнении H_0.

Рассмотрим процедуру и примеры проверки нулевой гипотезы для коэффициента корреляции на конкретном примере. Этот пример поможет показать логику и процедуру проверки статистических гипотез вообще. Взяты 10 наблюдений показателей инфляции и безработицы в США за 1931-1940 годы, для них рассчитан выборочный коэффициент корреляции, составивший -0,227. Связь отрицательная, что соответствует теории (кривая Филлипса), но значима ли она? Проверим гипотезу H_0: ρ=0 о равенстве нулю истинного значения коэффициента корреляции. Для проверки гипотезы H_0, как уже говорилось, следует использовать t-статистику с n-2 степенями свободы.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \tag{8}$$

Сравнивая определенное по выборочным данным значение статистики t с критическими точками, определяемыми по таблицам распределения Стьюдента, мы можем принять или отвергнуть нулевую гипотезу. В нашем примере t-статистика составляет -0,66. Зададим уровень значимости α=0,05, то есть 5%. Критическая (заштрихованная) область состоит из двух одинаковых "хвостов", площадь каждого из которых составляет 0,025. Рассмотрим таблицы вероятности того, что величина t-статистики превысит уровень z, то есть попадет в правый "хвост" распределения. Вероятность попасть только в правый "хвост", то есть в одностороннюю критическую область, равна $\alpha/2$, в нашем случае 0,025. Из таблицы найдем, что критическое значение z составляет 2,306. Это означает, что мы отвергли бы нулевую гипотезу только если $|t|>2{,}306$, а в нашем случае $|t|=0{,}66$. Итак, в нашем случае не исключается, что истинное значение коэффициента корреляции равно нулю, то есть на основе данной выборки не удалось сделать вывод о наличии статистически значимой линейной связи показателей инфляции и безработицы в США. Нельзя, впрочем, здесь сделать вывода и об отсутствии такой связи.

Вопросы к главе 15

1. Какой вид (аналитический и графический) имеют плотность распределения вероятности и функция распределения стандартного равномерного распределения, определенного на интервале $0 \le x \le 1$?
2. Какой вид (аналитический и графический) имеют плотность распределения вероятности и функция распределения стандартного нормального распределения?
3. В чем важность нормального распределения для экономического анализа?
4. Что такое распределение Стьюдента? Где и как оно применяется?
5. Что такое совместная, предельная и условная вероятности двух событий A и B? Каковы их определения и связь между ними?
6. Как определяется независимость событий?
7. Укажите основные вероятностные характеристики двух случайных величин и соотношения между ними.
8. Что такое ковариация и коэффициент корреляции двух случайных величин? Какое свойство случайных величин они характеризуют?
9. В каких случаях понятия некоррелированности и независимости двух случайных величин эквивалентны, а в каких различны?
10. Приведите пример совместной плотности распределения вероятности двух случайных величин и нарисуйте их линии уровня для различных значений коэффициента корреляции этих величин.
11. Как проверяется гипотеза о некоррелированности двух случайных величин?

ГЛАВА 16
МОДЕЛЬ ЛИНЕЙНОЙ РЕГРЕССИИ

16.1. Проблема оценивания линейной связи экономических переменных

Проблема изучения взаимосвязей экономических показателей является одной из важнейших проблем экономического анализа. Любая экономическая политика заключается в регулировании экономических переменных, и она должна основываться на знании того, как эти переменные влияют на другие переменные, являющиеся ключевыми для принимающего решения политика. Так, в рыночной экономике нельзя непосредственно регулировать темп инфляции, но на него можно воздействовать средствами бюджетно-налоговой и кредитно-денежной политики. Поэтому, в частности, должна быть изучена зависимость между предложением денег и уровнем цен. Невозможно строить, проверять или улучшать экономические модели без статистического анализа их переменных с использованием реальных статистических данных. Вся сфера экономических исследований может быть в определенном смысле охарактеризована как изучение взаимосвязей экономических переменных, и инструментарием их базового анализа являются методы статистики и эконометрики.

Изучение зависимостей экономических переменных начнем со случая двух переменных (обозначим их x и y). Этот случай наиболее прост и может быть рассмотрен графически. Предположим, что имеются ряды значений переменных, соответствующие им точки нанесены на график и соединены линией. Если это реальные статистические данные, то мы никогда не получим простую линию - линейную, квадратичную, экспоненциальную и т.д. Всегда будут присутствовать отклонения зависимой переменной, вызванные ошибками измерения, влиянием неучтенных величин или случайных факторов. Но если мы не получили, например, точную прямую линию, это еще не значит, что в основе рассматриваемой зависимости лежит нелинейная функция. Возможно, зависимость переменных линейна, и лишь случайные факторы приводят к некоторым отклонениям от нее. То же самое можно сказать и про любой другой вид функции. Связь переменных, на которую накладываются воздействия случайных факторов, называется статистической связью. Наличие такой связи заключается в том, что изменение одной переменной приводят к изменению математического ожидания другой переменной. Можно указать два типа взаимосвязей между переменными x и y. В одном случае может быть неизвестно, какая из двух переменных является независимой, и какая - зависимой. В этом случае переменные равноправны, и имеет смысл говорить о статистической взаимосвязи корреляционного типа. Оценка и анализ парной корреляции уже рассматривались в предыдущей главе. Другая ситуация возникает, если две исследуемые переменные не равноправны, но одна из них рассматривается как объясняющая (или независимая), а другая как объясняемая (или зависящая от первой). Если это так, то изменение одной из переменных служит причиной для изменения другой. Например, рост дохода ведет к увеличению потребления; снижение процентной ставки увеличивает инвестиции; увеличение валютного курса сокращает чистый экспорт. Это - тот случай, когда должно быть оценено уравнение регрессии $y=f(x)$. Уравнение регрессии - это формула статистической связи между переменными. Если эта формула линейна, то речь идет о линейной регрессии. Формула статистической связи двух переменных называется парной регрессией, зависимость от нескольких переменных - множественной регрессией. Например, Кейнсом была предложена линейная формула зависимости частного потребления C от располагаемого дохода Y_d: $C=C_0+b\ Y_d$, где $C_0>0$ - величина автономного потребления, $1>b>0$ - предельная склонность к потреблению.

Выбор формулы связи переменных называется спецификацией уравнения регрессии; в данном случае выбрана линейная формула. Однако до тех пор, пока не оценены количественные значения параметров C_0 и b, не проверена надежность сделанных оценок, эта формула остается лишь гипотезой. Оценка значений параметров выбранной формулы статистической связи переменных называется параметризацией уравнения регрессии. Как же оценить значения параметров и проверить надежность оценок? Рассмотрим вначале рис. 16.1.

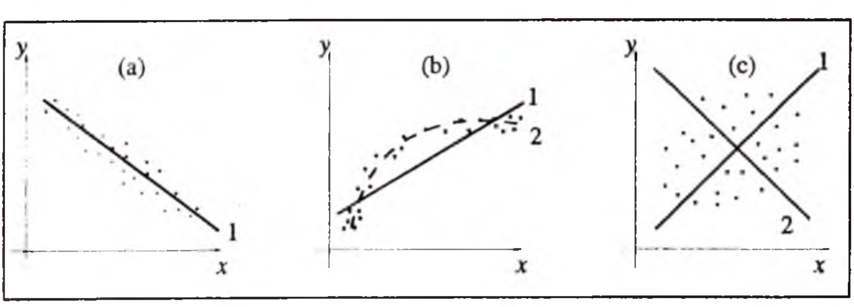


Рис. 16.1

На рисунке 16.1 изображены три ситуации:

- на графике (a) взаимосвязь x и y близка к линейной; прямая линия (1) здесь близка к точкам наблюдений, и последние отклоняются от нее лишь в результате сравнительно небольших случайных воздействий;
- на графике (b) реальная взаимосвязь величин x и y описывается нелинейной функцией (2), и какую бы мы ни провели прямую линию (например, 1), отклонения точек наблюдений от нее будут существенными и неслучайными;
- на графике (c) явная взаимосвязь между переменными x и y отсутствует; какую бы мы ни выбрали формулу связи, результаты ее параметризации будут здесь неудачными. В частности, прямые линии 1 и 2, проведенные через "центр" "облака" точек наблюдений и имеющие противоположный наклон, одинаково плохи для того, чтобы делать выводы об ожидаемых значениях переменной y по значениям переменной x.

16.2. Парная линейная регрессия. Метод наименьших квадратов

Начальным пунктом эконометрического анализа зависимостей обычно является оценка линейной зависимости переменных. Если имеется некоторое "облако" точек наблюдений, через него всегда можно попытаться провести такую прямую линию, которая является наилучшей в определенном смысле среди всех прямых линий, то есть "ближайшей" к точкам наблюдений по их совокупности. Для этого мы вначале должны определить понятие близости прямой к некоторому множеству точек на плоскости; меры такой близости могут быть различными. Однако любая разумная мера должна быть, очевидно, связана с расстояниями от точек наблюдений до рассматриваемой прямой линии (задаваемой уравнением $y = a + bx$).

Обычно в качестве критерия близости используется минимум суммы квадратов разностей наблюдений зависимой переменной y_i и теоретических, рассчитанных по уравнению регрессии значений $(a + bx_i)$:

$$Q = \sum_i e_i^2 = \sum_i (y_i - (a + bx_i))^2 \rightarrow \min. \tag{1}$$

Здесь считается, что y_i и x_i - известные данные наблюдений, a и b - неизвестные параметры линии регрессии. Поскольку функция Q непрерывна, выпукла и ограничена снизу нулем, она имеет минимум. Для соответствующих точке этого минимума значений a и b могут быть найдены простые и удобные формулы (они будут приведены ниже). Метод оценивания параметров линейной регрессии, минимизирующий сумму квадратов отклонений наблюдений зависимой переменной от искомой линейной функции, называется *Методом наименьших квадратов (МНК)*, или *Least Squares Method (LS)*.

"Наилучшая" по МНК прямая линия всегда существует, но даже наилучшая не всегда является достаточно хорошей. Если в действительности зависимость $y=f(x)$ является, например, квадратичной (как на рисунке 16.1(b)), то ее не сможет адекватно описать никакая линейная функция, хотя среди всех таких функций обязательно найдется "наилучшая". Если величины x и y вообще не связаны (рис. 16.1(c)), мы также всегда сможем найти "наилучшую" линейную функцию $y = a+bx$ для данной совокупности наблюдений, но в этом случае конкретные значения a и b определяются только случайными отклонениями переменных и сами будут очень сильно меняться для различных выборок из одной и той же генеральной совокупности. Возможно, на рис. 16.1(c) прямая 1 является наилучшей среди всех прямых линий (в смысле минимального значения функции Q), но любая другая прямая, проходящая через центральную точку "облака" (например, линия 2), ненамного в этом смысле хуже, чем прямая 1, и может стать наилучшей в результате небольшого изменения выборки.

Рассмотрим теперь задачу оценки коэффициентов парной линейной регрессии более формально. Предположим, что связь между x и y линейна: $y = \alpha+\beta x$. Здесь имеется в виду связь между всеми возможными значениями величин x и y, то есть для генеральной совокупности. Наличие случайных отклонений, вызванных воздействием на переменную y множества других, неучтенных в нашем уравнении факторов и ошибок измерения, приведет к тому, что связь наблюдаемых величин x_i и y_i приобретет вид $y_i = \alpha + \beta x_i + \epsilon_i$. Здесь ϵ_i - случайные ошибки (отклонения, возмущения). Задача состоит в следующем: по имеющимся данным наблюдений $\{x_i\}$, $\{y_i\}$ оценить значения параметров α и в, обеспечивающие минимум величины Q. Если бы были известны точные значения отклонений ϵ_i, то можно было бы (в случае правильности предполагаемой линейной формулы) рассчитать значения параметров α и β. Однако значения случайных отклонений в выборке неизвестны, и по наблюдениям x_i и y_i можно получить оценки параметров α и β, которые сами являются случайными величинами, поскольку соответствуют случайной выборке. Пусть a - оценка параметра α, b - оценка параметра β. Тогда оцененное уравнение регрессии будет иметь вид: $y_i=a+bx_i+e_i$, где e_i - наблюдаемые значения ошибок ϵ_i.

Для оценки параметров α и в воспользуемся МНК, который минимизирует сумму квадратов отклонений фактических значений y_i от расчетных (см. (16.1)). Минимум ищется по переменным a и b.

При использовании МНК к ошибкам ε_i предъявляются следующие требования, называемые условиями Гаусса–Маркова:

1) величина $\in_i$ является случайной переменной;

2) математическое ожидание $\in_i$ равно нулю: $M(\in_i) = 0$;

3) дисперсия $\in_i$ постоянна: $D(\in_i) = D(\in_j) = \sigma^2$ для всех i, j;

4) значения $\in_i$ независимы между собой. Откуда вытекает, в частности, что

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0 & \text{при } i \neq j, \\ \sigma^2 & \text{при } i = j. \end{cases} \tag{2}$$

5) величины $\in_i$ статистически независимы со значениями x_i.

Известно, что, если условия 1)-5) выполняются, то оценки, сделанные с помощью МНК, обладают следующими свойствами:

1) Оценки являются несмещенными, т.е. математическое ожидание оценки каждого параметра равно его истинному значению: $M(a) = \alpha$; $M(b) = \beta$. Это вытекает из того, что $M(\in_i) = 0$, и говорит об отсутствии систематической ошибки в определении положения линии регрессии.

2) Оценки состоятельны, так как дисперсия оценок параметров при возрастании числа наблюдений стремится к нулю: $\lim\limits_{n\to\infty} D(a) = 0$; $\lim\limits_{n\to\infty} D(b) = 0$. Иначе говоря, если n достаточно велико, то практически наверняка a близко к α, а b близко к β: надежность оценки при увеличении выборки растет.

3) Оценки эффективны, они имеют наименьшую дисперсию по сравнению с любыми другими оценками данного параметра, линейными относительно величин y_i. В англоязычной литературе такие оценки называются BLUE (*Best Linear Unbiased Estimators* - наилучшие линейные несмещенные оценки).

Перечисленные свойства не зависят от конкретного вида распределения величин $\in_i$, тем не менее обычно предполагается, что они распределены нормально $N(0; \sigma^2)$. Эта предпосылка необходима для проверки статистической значимости сделанных оценок и определения для них доверительных интервалов. При ее выполнении оценки МНК имеют наименьшую дисперсию не только среди линейных, но среди всех несмещенных оценок.

Если предположения 3) и 4) нарушены, то есть дисперсия возмущений непостоянна и/или значения $\in_i$ связаны друг с другом, то свойства несмещенности и состоятельности сохраняются, но свойство эффективности - нет.

При невыполнении предположения 5) может нарушаться и свойство несмещенности оценок, являющееся наиболее важным в эконометрическом анализе. Значительная часть современной эконометрической теории посвящена анализу выполнения данного свойства (в совокупности с остальными) в различных конкретных ситуациях, а также выяснению и корректировке последствий его невыполнения.

Рассмотрим теперь процедуру оценивания параметров парной линейной регрессии a и b. Для того, чтобы функция $Q = \sum_i e_i^2 = \sum_i (y_i - (a + bx_i))^2$ достигала минимума, необходимо равенство нулю ее частных производных:

$$\begin{cases} Q'_a = -2\sum_i (y_i - a - bx_i) = 0 \\ Q'_b = -2\sum_i (y_i - a - bx_i)x_i = 0 \end{cases} \Rightarrow \begin{cases} \sum_i y_i - na - b\sum_i x_i = 0 & (3) \\ \sum_i y_i x_i - a\sum_i x_i - b\sum_i x_i^2 = 0 & (4) \end{cases}$$

Если уравнение (16.3) разделить на n, то получим $\bar{y} = a + b\bar{x}$ (здесь $\bar{y} = \frac{\sum_i y_i}{n}$; $\bar{x} = \frac{\sum_i x_i}{n}$ - средние значения x и y). Таким образом, линия регрессии проходит через точку со средними значениями x и y. Подставив величину a из (16.3) в (16.4), получаем

$$\sum_i y_i x_i = \sum_i x_i(\bar{y} - b\bar{x}) + b\sum_i x_i^2 = n\bar{x}(\bar{y} - b\bar{x}) + b\sum_i x_i^2.$$

Откуда

$$b = \frac{\sum_i y_i x_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}; \qquad (5)$$

$$a = \bar{y} - b\bar{x}. \qquad (6)$$

Иначе можно записать, что $b = \frac{\operatorname{cov}(x,y)}{D(x)} = r\frac{\sqrt{D(y)}}{\sqrt{D(x)}}$ (где r - коэффициент корреляции x и y). Таким образом, коэффициент регрессии пропорционален показателю ковариации и коэффициенту корреляции x и y, а коэффициенты этой пропорциональности служат для соизмерения перечисленных разноразмерных величин. Оценки a и b, очевидно, являются линейными относительно y_i (если x_i считать коэффициентами) - выше об этом упоминалось.

Итак, если коэффициент r уже рассчитан, то легко рассчитать коэффициент парной регрессии, не решая системы уравнений. Ясно также, что если рассчитаны линейные регрессии $x(y)$ и $y(x)$, то произведение коэффициентов b_x и b_y равно r^2:

$$b_x b_y = r\frac{\sqrt{D(y)}}{\sqrt{D(x)}} \cdot r\frac{\sqrt{D(x)}}{\sqrt{D(y)}} = r^2. \qquad (7)$$

16.3. Анализ статистической значимости коэффициентов линейной регрессии

Величины y_i, соответствующие данным x_i при некоторых теоретических значениях α и β, являются случайными. Следовательно, случайными являются и рассчитанные по ним значения коэффициентов a и b. Их математические ожидания при выполнении предпосылок об отклонениях ϵ_i равны, соответственно, α и β. При этом оценки тем надежнее, чем меньше их разброс вокруг α и β, то есть дисперсия. По определению дисперсии $D(b) = M(b\text{-}\beta)^2$; $D(a) = M(a\text{-}\alpha)^2$. Надежность получаемых оценок a и b зависит, очевидно, от дисперсии случайных отклонений ϵ_i, но поскольку по данным выборки эти отклонения (и, соответственно, их дисперсия) оценены быть не могут, они заменяются при анализе надежности оценок коэффициентов регрессии на отклонения переменной y от оцененной линии регрессии $e_i = y_i - a - bx_i$.

Можно доказать (доказательство опускаем), что $D(b) = S_b^2 = \dfrac{S^2}{\sum\limits_i (x_i - \bar{x})^2}$; $D(a) = S_a^2 = \dfrac{S^2 \sum\limits_i x_i^2}{n\sum\limits_i (x_i - \bar{x})^2}$, где $S^2 = \dfrac{\sum\limits_i e_i^2}{n - 2}$ - мера разброса зависимой переменной вокруг линии регрессии (*необъясненная дисперсия*). S_a и S_b - стандартные отклонения случайных величин a и b. Полученный результат можно проинтерпретировать следующим образом.

Коэффициент b есть мера наклона линии регрессии. Очевидно, чем больше разброс значений y вокруг линии регрессии, тем больше (в среднем) ошибка в определении наклона линии регрессии. Если такого разброса нет совсем ($\epsilon_i = 0$ и, следовательно, $\sigma^2=0$), то прямая определяется однозначно и ошибки в расчете коэффициентов a и b отсутствуют (а отсюда и значение S^2, "замещающее" σ^2, равно нулю).

На рис. 16.2a отклонения в значениях переменной y от линии регрессии отсутствуют, и через три точки проводится та же прямая, что и через любые две из них. На рис. 16.2b через три точки проводится такая же линия регрессии, но колебания значений переменной у вокруг этой линии значительны. Поэтому через пары точек (1,2) и (1,3) проходят совершенно разные прямые, отличные от общей прямой. Следовательно, стандартные ошибки коэффициентов регрессии в этом случае будут значительными.

В знаменателе величины $D(b)$ стоит сумма квадратов отклонений x_i от среднего значения $\bar{x}$. Эта сумма велика в том случае, если регрессия оценена на достаточно широком диапазоне значений переменной x, и в этом случае, при данном уровне разброса y_i, очевид-

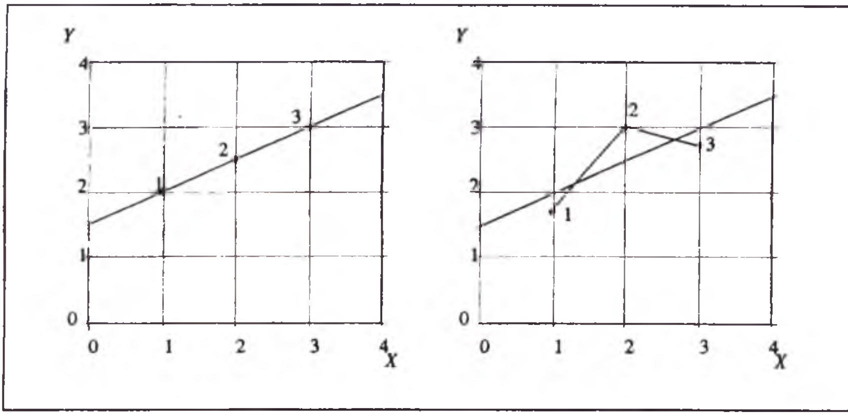


Рис. 16.2a *Рис. 16.2b*

но, ошибка в оценке величины наклона прямой будет меньше, чем при малом диапазоне изменения переменной x. Попробуйте провести прямую по двум точкам: если x_1 и x_2 лежат рядом, то даже небольшое изменение одного из y_i существенно меняет наклон прямой (если x_1 и x_2 далеки друг от друга - ситуация обратная).

Так, на рисунке 16.3 через пары точек (1,2) и (1a,2) проходят одни и те же прямые, в то же время разброс переменной x для первой из пар больше. Если у второй точки из каждой пары изменить значение переменной y (перевести ее в точку 2a или 2b), то наклон прямой для пары (1,2) изменится значительно меньше, чем для пары (1a,2).

Кроме того, чем больше (при прочих равных) число наблюдений n, тем больше $\sum_i (x_i - \bar{x})^2$ и, тем самым, меньше стандартная ошибка оценки. Дисперсия свободного члена уравнения регрессии равна $D(a) = D(b)\frac{\sum_i x_i^2}{n}$ - она пропорциональна $D(b)$ и, тем самым, также соответствует уже сделанным пояснениям о влиянии разброса y_i вокруг регрессионной прямой и разброса x_i на стандартную ошибку. Чем сильнее меняется наклон прямой, проведенной через данную точку $(\bar{x},\bar{y})$, тем больше разброс значений свободного члена, характеризующего точку пересечения этой прямой с осью y. Кроме того, дисперсия и стандартная ошибка свободного члена тем больше, чем больше средняя величина x_i^2. При больших по модулю значениях x даже небольшое изменение наклона регрессионной прямой может вызвать большое изменение оценки свободного члена, поскольку в этом случае в среднем велико расстояние от точек наблюдений до оси y.

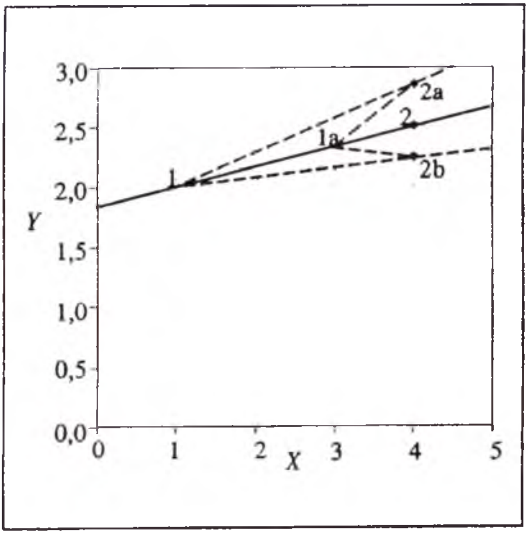


Рис. 16.3

На рис. 16.4 через пары точек (1,2) и (3,4) проходит одна и та же прямая линия. Ее свободный член равен a. Для второй из этих пар значения переменной x больше по абсолютной величине (при одинаковом разбросе значений x и y). Если в первой из этих пар от точки 1 перейти к точке 1а, а во второй - от точки 3 к 3а, что вызвано одинаковыми изменениями одного из значений переменной y, то обе линии становятся горизонтальными. Изменения коэффициента наклона прямой одинаковы, но свободный член в первом случае становится равным a_l, а во втором - a_2, - таким образом, он меняется значительно больше там, где больше абсолютные значения переменной x.

Формально значимость оцененного коэффициента регрессии b может быть проверена с помощью анализа его отношения к своему стандартному отклонению $S_b = \sqrt{D(b)}$. Эта величина в случае выполнения исходных предпосылок модели имеет t-распределение Стьюдента с (n-2) степенями свободы (n - число наблюдений). Она называется *t-статистикой*:

$$t = \frac{b}{\sqrt{D(b)}} = \frac{b}{S_b}. \qquad (8)$$

Для t-статистики проверяется нулевая гипотеза, то есть гипотеза о равенстве ее нулю. Очевидно, $t = 0$ равнозначно $b = 0$, поскольку t пропорциональна b.

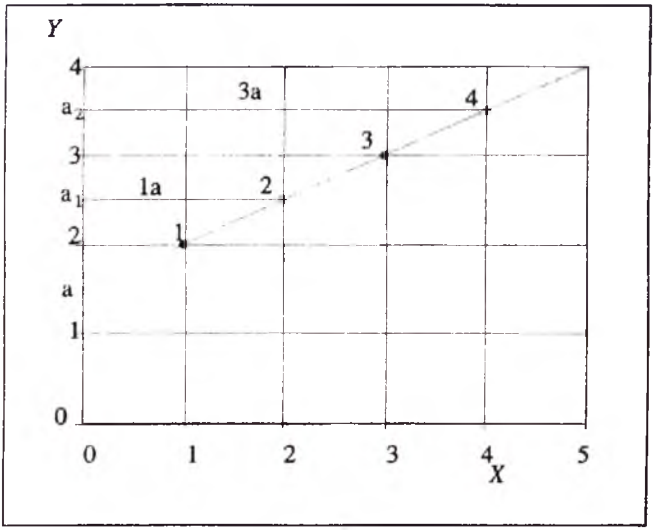


Рис. 16.4

Рассмотрим конкретный пример, уже затрагивавшийся в предыдущей главе. Пусть *INF* - темп инфляции, *U* - уровень безработицы в США в 1931 - 1940 годы (10 наблюдений). Точки наблюдений показаны на рис. 16.5.

Из рис.16.5 можно видеть, что, возможно, есть некоторая отрицательная связь показателей *INF* и *U*, но вряд ли этот рисунок подтверждает наличие статистически значимой линейной связи. Для проверки этого вывода оценена парная регрессия $INF = 5{,}07 - 0{,}32 \cdot U$. Оценена величина

$$S_b^2 = \frac{S^2}{\sum\limits_i x_i^2 - \bar{x}^2} = \frac{\sum\limits_i e_i^2}{(n-2)\left(\sum\limits_i x_i^2 - \bar{x}^2\right)} = \frac{\sum\limits_i (y_i - a - bx_i)^2}{8\left(\sum\limits_i x_i^2 - \bar{x}^2\right)} \approx 0{,}236 \Rightarrow$$

$S_b \approx 0{,}486.$

Отсюда $t = -\dfrac{0{,}32}{0{,}486} \approx -0{,}658$. Зададим уровень значимости 0,1 при двусторонней альтернативной гипотезе (то есть если величина $b \neq 0$, то она может быть как положительной, так и отрицательной). Таблицы для *t*-статистик обычно публикуются для односторонней альтернативной гипотезы ($t>0$), поэтому найдем критическое значение для уровня значимости 0,05 (доверительная вероятность 0,95) с $(n-2)=8$ степенями свободы $t_{8;0,95}=1{,}860$ и сравним с ним $|t|=0{,}658$.

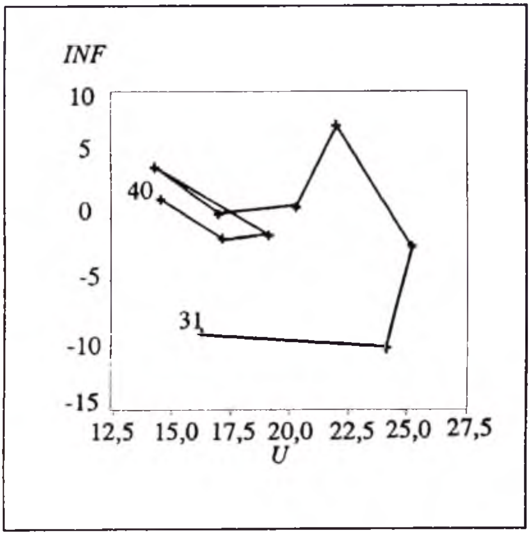


Рисунок 16.5. Уровень инфляции (*INF*) и безработицы (*U*) в США в 1931-1940 гг.

Поскольку $|t|<1{,}860$, нулевая гипотеза $\{t = 0\}$ не может быть отвергнута при заданном уровне значимости. Иными словами, нельзя считать (грубо говоря), что уровень инфляции в рассматриваемый период значимо зависел от показателя безработицы. Если уровень значимости задать равным 0,3, то $t_{8;0,85} = 1{,}108 > 0{,}658$, - даже при такой слабой значимости нулевая гипотеза не может быть отвергнута.

Проверка значимости коэффициента парной линейной регрессии эквивалентна проверке значимости коэффициента корреляции переменных x и y. В этом можно убедиться, сравнив значения t-статистик для коэффициента корреляции в предыдущей главе и коэффициента регрессии b (пример рассматривается один и тот же). Эти значения одинаковы и равны -0,658. Соответственно, и уровень значимости у них одинаков.

При оценке значимости коэффициента линейной регрессии можно использовать следующее грубое правило. Если стандартная ошибка коэффициента больше его модуля ($t < 1$), то он не может быть признан хорошим (значимым), поскольку доверительная вероятность здесь при двусторонней альтернативной гипотезе составляет лишь менее, чем приблизительно 0,7. Если стандартная ошибка меньше модуля коэффициента, но больше его половины ($1 < t < 2$), то сделанная оценка может рассматриваться как более или менее значимая. Доверительная вероятность здесь примерно от 0,7 до 0,95. Значение t от 2 до 3 свидетельствует о весьма значимой связи (дове-

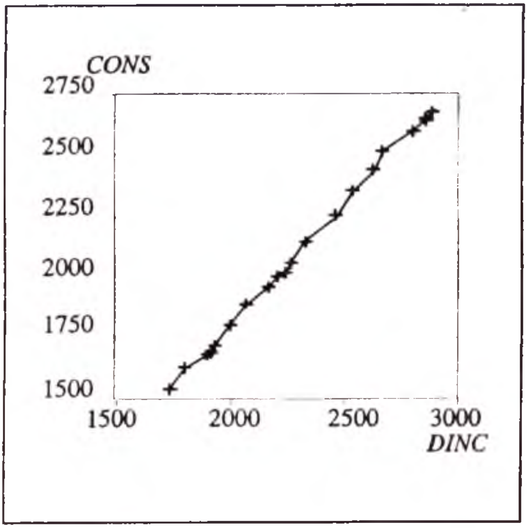


Рисунок 16.6. Объем частного потребления (*CONS*, *C*) и располагаемого дохода (*DINC*, Y_d) в США в 1971-1990 гг. (млрд.долл., 1982 г.)

рительная вероятность от 0,95 до 0,99), и $t > 3$ есть практически стопроцентное свидетельство ее наличия. Конечно, в каждом случае играет роль число наблюдений; чем их больше, тем надежнее при прочих равных выводы о наличии связи и тем меньше верхняя граница доверительного интервала для данных числа степеней свободы и уровня значимости. Однако эти различия существенны лишь для малых n, а при n порядка 10 и более сформулированные правила приблизительно верны.

Для иллюстрации действительно значимой линейной связи показателей рассмотрим величины частного потребления и располагаемого дохода в США за 1971-1990 годы. Динамика этих показателей показана на рис. 16.6.

На рисунке 16.6 явно просматривается четкая линейная зависимость объема частного потребления от величины располагаемого дохода. Уравнение парной линейной регрессии, оцененное по этим данным, имеет вид: $C = -217{,}6 + 1{,}007 \cdot Y_d$. Стандартные ошибки для свободного члена и коэффициента парной регрессии равны, соответственно, 28,4 и 0,012, а t-статистики - -7,7 и 81,9. Обе они по модулю существенно превышают 3, следовательно, их статистическая значимость весьма высока. Впрочем, несмотря на то, что здесь удалось оценить статистически значимую линейную функцию потребления, в ней нарушены сразу две предпосылки Кейнса - уровень автономного потребления C_0 оказался отрицательным, а предельная склонность к потреблению превысила единицу. Очевидно, в рассматриваемый период наблюдался процесс “вытеснения” потреблением некоторых других составляющих ВНП (в частности - чистого экспорта).

16.4. Сравнение истинных и оцененных зависимостей

Соотношение между истинной зависимостью между переменными (в генеральной совокупности) и зависимостью, оцененной по выборочным данным проще всего показать на примере соотношения между доходами и расходами. Пусть, к примеру, в небольшом городке проживают сто семей (генеральная совокупность), доходы которых (X_k) можно отнести к одной из пяти групп ($k = 1,...,5$). Предположим также для простоты, что распределение людей по доходам - равномерное, то есть в каждую группу входят 20 семей. Собрав данные по расходам на члена семьи, нанесем их в виде точек на график, по вертикальной оси которого отложим расходы, а по горизонтальной - доходы.

На рис. 16.7 видно, что, во-первых, даже внутри группы с одним доходом расходы людей различны, что объясняется различием вкусов, потребностей, количеством членов в семье и другими факторами, которые не входят в число переменных, объясняющих расходы, и представляемыми в виде случайного (по отношению к доходам) компонента расходов. Во-вторых, можно заметить, что, в среднем, расходы растут с увеличением доходов.

Обозначая средние по k-й группе дохода (в генеральной совокупности) расходы $M[Y|X_k]$, можно представить тенденцию увеличения расходов с доходами в виде положительной линейной зависимости

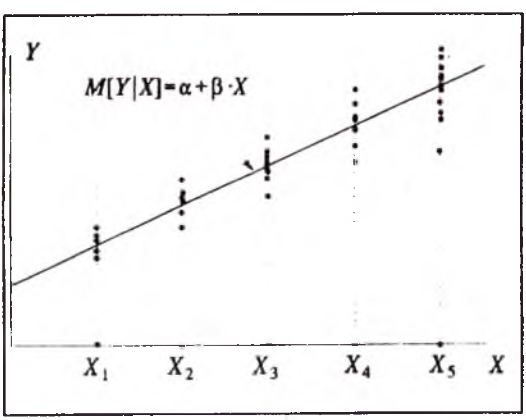


Рис. 16.7

$$M[\,Y|X] = \alpha + \beta \cdot X, \tag{9}$$

которая предполагается истинной зависимостью между средними расходами и доходами. Для неусредненных расходов в эту зависимость следует добавить случайный член ϵ, описывающий разброс расходов внутри группы с одним доходом, обусловленный действием всех остальных факторов, кроме доходов.

$$Y = \alpha + \beta \cdot X + \epsilon. \tag{10}$$

Эта зависимость предполагается истинной зависимостью между индивидуальными расходами и доходами (в генеральной совокупности).

Теперь обратимся к выборочным данным о расходах, собранным путем выборочного опроса части жителей городка. Считая выборку репрезентативной, предположим, для простоты, она включает по одному человеку из каждой группы дохода. Отображая выборочные точки на графике, мы можем провести через них линию регрессии, соответствующую уравнению $Y = a + b \cdot X$, коэффициенты a и b в котором рассчитываются по обычным формулам линейной регрессии. Если учесть, что наблюдаемые значения Y_k не лежат на линии регрессии $(a+b \cdot X_k)$, то в это уравнение надо добавить выборочные случайные возмущения e $(e_k = Y_k - a - b \cdot X_k)$, являющиеся аналогами случайных возмущений ϵ в генеральной совокупности:

$$Y_k = \mathrm{a} + b \cdot X_k + e_k. \tag{11}$$

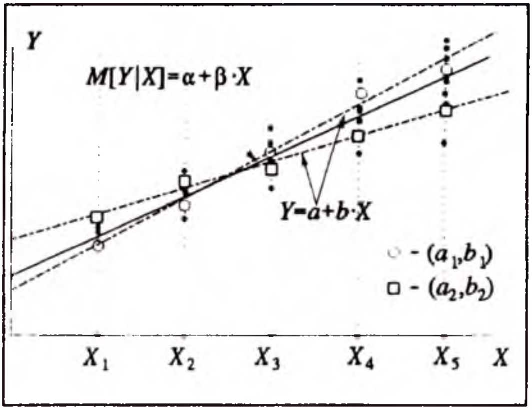


Рис. 16.8

Таким образом, мы имеем две линейных регрессии: одну для генеральной совокупности, коэффициенты в которой обычно обозначаются греческими буквами, и другую для выборки, коэффициенты в которой обычно обозначаются латинскими буквами. Коэффициенты линейной зависимости для генеральной совокупности нам неизвестны, и мы должны их оценить, пользуясь выборочными данными. Коэффициенты выборочной линейной регрессии a и b являются выборочными оценками коэффициентов α и β в генеральной совокупности.

Из рис. 16.8 видно, что выборочные линии регрессии имеют разный наклон и разные точки пересечения с осью Y для различных выборок. Более того, при положительном наклоне генеральной регрессии наклон выборочной линии регрессии может оказаться для некоторых выборок отрицательным, что, однако, не будет свидетельствовать об истинной отрицательной связи исследуемых величин. Для того чтобы убедиться в этом, следует помимо коэффициентов регрессии находить их стандартные отклонения и t-статистики, по которым можно судить о статистической значимости полученных выборочных коэффициентов регрессии.

16.5. Множественная линейная регрессия

Значения экономических переменных определяются обычно влиянием не одного, а нескольких объясняющих факторов. В таком случае зависимость $y = f(x)$ означает, что x - вектор, содержащий m компонентов: $x = (x_1, x_2, \dots, x_m)$. Задача оценки статистической взаимосвязи переменных y и $x = (x_1, x_2, \dots, x_m)$ формулируется аналогично случаю парной регрессии. Записывается функция $y = f(\alpha, x) + \epsilon$, где α - вектор параметров, ϵ - случайная ошибка. Предполагается, что эта функция связывает переменную y с вектором независимых переменных x для данных генеральной совокупности. Как и в случае парной регрессии, предполагается, что ошибки ϵ_i являются случайными величинами с нулевым математическим ожиданием и постоянной дисперсией; ϵ_i и ϵ_j статистически независимы при $i j$. Кроме того, для проверки статистической значимости оценок α обычно предполагается, что ошибки ϵ_i нормально распределены. По данным наблюдений выборки размерности n требуется оценить значения параметров α, то есть провести параметризацию выбранной формулы (спецификации) зависимости.

Мы будем говорить о линейной зависимости y от x, то есть о множественной линейной регрессии. Теоретическое уравнение регрессии имеет вид:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + \epsilon. \tag{12}$$

Здесь α - вектор неизвестных параметров размерности $(m + 1)$. Пусть имеется n наблюдений вектора x и зависимой переменной y. Для того, чтобы формально можно было решить задачу, то есть найти некоторый наилучший вектор параметров, должно быть $n \geq m+1$. Если это условие не выполняется, то можно найти бесконечно много разных векторов коэффициентов, при которых линейная формула связывает между собой x и y для имеющихся наблюдений абсолютно точно. Если, в частном случае, $n = m+1$ (например, при двух объясняющих переменных в уравнении $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$ и трех наблюдениях), то оценки коэффициентов α рассчитываются единственным образом - путем решения системы линейных уравнений $\{y_j=\alpha_0+\alpha_1x_{1j}+\alpha_2x_{2j}+...+\alpha_mx_{mj};\ j=1,2,...,n$ - индекс наблюдения$\}$. Так, через три точки-наблюдения в трехмерном пространстве можно провести единственную плоскость определяемую параметрами $\alpha_0,\alpha_1,\alpha_2$. Если число наблюдений больше минимально необходимого, то есть $n > m+1$, то уже нельзя подобрать линейную формулу, в точности удовлетворяющую всем наблюдениям, и возникает необходимость оптимизации, то есть выбора наилучшей формулы-приближения для имеющихся наблюдений. Положительная разность (n-m-1) в этом случае называется числом степеней свободы. Если число степеней свободы мало, то статистическая надежность оцениваемой формулы невысока. Так, если проведена плоскость "в точности" через имеющиеся три точки наблюдений, любая четвертая точка-наблюдение из той же генеральной совокупности будет практически наверняка лежать вне этой плоскости, возможно - достаточно далеко от нее. Обычно при оценке множественной регрессии для обеспечения статистической надежности требуется, чтобы число наблюдений по крайней мере в 3 раза превосходило число оцениваемых параметров.

Задача построения множественной линейной регрессии состоит в нахождении (m+1)-мерного вектора а, элементы которого есть оценки соответствующих элементов вектора α. Критерии оценивания, как и в случае парной регрессии, могут быть различными; мы будем вновь использовать метод наименьших квадратов (МНК).

Уравнение регрессии с оцененными параметрами имеет вид

$$y = a_0 + a_1x_1 + a_2x_2 + ... + a_mx_m + e, \tag{13}$$

и критерием для нахождения вектора (a) является $\min\sum_i e_i^2$. Оцененное уравнение должно описать как общий тренд (тенденцию) изменения зависимой переменной y, так и отклонения от этого тренда. Проблема здесь состоит не только в том, чтобы объяснить возможно большую долю колебаний переменной y, но и отделить влияние каждого из факторов, рассматриваемых как объясняющие переменные.

При выполнении предпосылок 1)-4) относительно ошибок e_i оценки параметров множественной линейной регрессии являются несмещенными, состоятельными и эффективными. Отклонение зависимой переменной y в j-м наблюдении от линии регрессии, e_j, записывается следующим образом: $e_j = y_j - a_0 - a_1x_{j1} - a_2x_{j2} - ... - a_mx_{jm}$. Обозначим сумму квадратов этих величин, которую нужно минимизировать в соответствии с методом наименьших квадратов, через Q:

$$Q = \sum_j e_j^2 = \sum_j (y_j - (a_0 + \sum_i a_ix_{ji}))^2 \to \min. \tag{14}$$

Минимизируемая функция Q является квадратичной относительно неизвестных величин a_i. Необходимым условием ее минимума является равенство нулю всех ее частных производных по a_i. Частные производные квадратичной функции являются линейными функциями, и, приравнивая их всех к нулю, мы получим систему из $(m+1)$ линейных уравнений с $(m+1)$ неизвестными. Такая система имеет обычно единственное решение (за исключением особого случая, когда столбцы ее линейно зависимы и решения нет или их бесконечно много; однако данные реальных статистических наблюдений к такому особому случаю, вообще говоря, никогда не приводят). Данная система называется системой нормальных уравнений. Ее решение в явном виде удобнее всего выписать в векторно-матричной форме, иначе оно становится слишком громоздким. Векторно-матричная запись и вывод решения системы нормальных уравнений приведены в Приложении; при начальном ознакомлении с проблемой оно может быть опущено.

Для анализа статистической значимости полученных коэффициентов множественной линейной регрессии необходимо, как и в случае парной регрессии, оценить дисперсию и стандартные отклонения коэффициентов a_j.

В случае парной регрессии $D(b) = S_b^2 = \dfrac{\sum_i e_i^2}{(n-2)\sum_i (x_i - \bar{x})^2}$; в общем случае $D(a_j) = \dfrac{\sum_i e_i^2}{(n-m-1)} \cdot Z_{jj}$, (где Z_{jj} - диагональный элемент матрицы $(X^TX)^{-1}$ - см. Приложение). Соответственно, стандартное отклонение $S_{aj} = \sqrt{D(a_j)}$, и для проверки нулевой гипотезы для каждого из коэффициентов a_j рассчитываются, как и в случае парной регрессии, t-статистики: $t = \dfrac{a_j}{S_{aj}}$, имеющие распределение Стьюдента с $(n\text{-}m\text{-}1)$ степенями свободы.

Если $(n\text{-}m\text{-}1)$, то есть число степеней свободы, достаточно велико (не менее 8 - 10), то при 5%-ном уровне значимости и двусторонней альтернативной гипотезе критическое значение t-статистики приблизительно равно двум. Здесь, как и в случае парной регрессии, можно приближенно считать оценку незначимой, если t-статистика по модулю меньше единицы, и весьма надежной, если модуль t-статистики больше трех. Другие критерии качества полученного уравнения регрессии будут рассмотрены в следующей главе. Там же будут приведены и примеры статистического анализа значимости коэффициентов множественной линейной регрессии.

ПРИЛОЖЕНИЕ

Расчет вектора коэффициентов множественной линейной регрессии

Пусть $e_i = y_i - a_0 - a_1x_{i1} - a_2x_{i2} - \ldots - a_mx_{im}$, где i - индекс наблюдения. Сумма квадратов отклонений e_i может быть записана как произведение вектор-строки

$\{e_i\} = e^T$ на вектор-столбец $\{e_i\} = e$ (e^T - вектор-столбец, транспонированный в строку). Вектор-столбец e, в свою очередь, может быть записан как $e = y - Xa$, где y - вектор-столбец наблюдений зависимой переменной, X - матрица n $(m+1)$, в которой каждая из n строк представляет наблюдение вектора значений независимых переменных x_i:

$$x = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1m} \\ 1 & x_{21} & x_{22} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix};$$

a - вектор-столбец $(a_0, a_1, \ldots, a_m)$.

Отсюда $Q = e^Te = (y-Xa)^T(y-Xa) = y^Ty - a^TX^Ty - y^TXa + a^TX^TXa = y^Ty - 2a^TX^Ty + a^TX^TXa$. Мы воспользовались здесь тем, что $(y - Xa)^T = y^T - (Xa)^T$; $(Xa)^T = a^TX^T$; $a^TX^Ty = y^TXa$.

Все эти свойства легко проверить, расписав поэлементно все матрицы и выполнив с ними нужные действия.

Теперь нужно записать необходимые условия экстремума выражения Q. Оно состоит в равенстве нулю всех частных производных $\frac{\partial Q}{\partial a_i}$. Вектор $\frac{\partial Q}{\partial a}$ можно записать компактно как $\frac{\partial Q}{\partial a} = -2X^Ty + 2(X^TX)a$. Это можно показать следующим образом: пусть $(X^TX) = X'$ - матрица $(m+1)$ $(m+1)$;

$$Q_1 = a^TX'a = \left[(a_0, a_1, \ldots, a_m) \cdot \begin{bmatrix} x'_{11} & x'_{12} & \ldots & x'_{1,m+1} \\ x'_{21} & x'_{22} & \ldots & x'_{2,m+1} \\ \ldots & \ldots & \ldots & \ldots \\ x'_{m+1,1} & x'_{m+1,2} & \ldots & x'_{m+1,m+1} \end{bmatrix}\right] \cdot \begin{bmatrix} a_0 \\ a_1 \\ \ldots \\ a_m \end{bmatrix} =$$

$$= \left(\sum_j a_j x'_{j1} + \sum_j a_j x'_{j2} + \ldots + \sum_j a_j x'_{j,m+1}\right)^T (a) = \sum_i a_i \sum_j a_j x'_{ji} = \sum_{ij} a_i a_j x'_{ij}.$$

Отсюда легко видно, что $\frac{\partial Q_1}{\partial a_i} = 2\sum_j a_j x_{ij}'$, то есть $\frac{\partial Q_1}{\partial a} = 2(X^TX)a$.

Ясно также, что если обозначить вектор $X^Ty = y'$, то $a^Ty' = \sum_j a_j y_j'$ и

$$\frac{\partial(a^Ty')}{\partial a} = y' = X^Ty.$$

Поскольку $Q = y^Ty\text{-}2a^Ty'+Q_1$, и $\frac{\partial y^Ty}{\partial a}=0$ (так как y^Ty - константа),

$\frac{\partial Q}{\partial a} = 0 \Rightarrow -2X^Ty + 2(X^TX)a = 0 \Rightarrow X^Ty = X^TXa$, откуда

$$\mathbf{a = (X^TX)^{-1} X^Ty.}$$

Здесь $(X^TX)^{-1}$ - матрица, обратная к (X^TX), то есть такая, которая при умножении на матрицу (X^TX) дает единичную матрицу. Таким образом, мы получили формулу расчета вектора коэффициентов регрессии в векторно-матричной записи.

Вопросы к главе 16

1. Что такое линейная регрессия?
2. Что такое спецификация и параметризация уравнения регрессии? Как они осуществляются?
3. Какими могут быть критерии качества оценки линейной регрессии?
4. В чем сущность метода наименьших квадратов (МНК)?
5. Сформулируйте общую задачу статистической оценки параметров на примере оценки параметров линейной регрессии.
6. Каковы предпосылки о свойствах отклонений зависимой переменной от теоретической линии регрессии?
7. Сформулируйте свойства несмещенности, состоятельности и эффективности оценок параметров. Обладают ли этими свойствами оценки параметров линейной регрессии, полученные с помощью МНК?
8. В чем различие, смысловое и количественное, теоретических значений коэффициентов регрессии α и β и их оценок a и b?
9. Какие факторы влияют на величину стандартных ошибок коэффициентов a и b?
10. Как связан коэффициент регрессии b с коэффициентом корреляции величин x и y?
11. Имеют ли коэффициенты a и b размерность?
12. Какой показатель характеризует долю объясненной с помощью регрессии дисперсии в общей дисперсии зависимой переменной?
13. Каким образом проверяется нулевая гипотеза для коэффициента регрессии b?
14. Стандартная ошибка коэффициента b равна $b/2$. Можно ли в этом случае говорить о наличии зависимости y от x? Если можно, то что именно?

ГЛАВА 17
ЛИНЕЙНАЯ РЕГРЕССИЯ: СТАТИСТИЧЕСКИЙ АНАЛИЗ МОДЕЛИ. ПРОГНОЗИРОВАНИЕ

В главе 16 рассмотрена постановка задачи оценивания уравнения линейной регрессии, показан способ ее решения. Однако оценка параметров конкретного уравнения является лишь отдельным этапом длительного и сложного процесса построения эконометрической модели. Первое же оцененное уравнение очень редко является удовлетворительным во всех отношениях. Обычно приходится постепенно подбирать формулу связи и состав объясняющих переменных, анализируя на каждом этапе качество оцененной зависимости. Этот анализ качества включает статистическую и содержательную составляющую. Проверка статистического качества оцененного уравнения состоит из следующих элементов:

- проверка статистической значимости каждого коэффициента уравнения регрессии;
- проверка общего качества уравнения регрессии;
- проверка свойств данных, выполнение которых предполагалось при оценивании уравнения.

Под содержательной составляющей анализа качества понимается рассмотрение экономического смысла оцененного уравнения регрессии: действительно ли значимыми оказались объясняющие факторы, важные с точки зрения теории; положительны или отрицательны коэффициенты, показывающие направление воздействия этих факторов; попали ли оценки коэффициентов регрессии в предполагаемые из теоретических соображений интервалы.

Методика проверки статистической значимости каждого отдельного коэффициента уравнения линейной регрессии была рассмотрена в предыдущей главе. Перейдем теперь к другим этапам проверки качества уравнения.

17.1. Проверка общего качества уравнения регрессии. Коэффициент детерминации R^2

Для анализа общего качества оцененной линейной регрессии используют обычно коэффициент детерминации R^2, называемый также квадратом коэффициента множественной корреляции. Для случая парной регрессии это квадрат коэффициента корреляции переменных x и y. Коэффициент детерминации рассчитывается по формуле

$$R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2} \tag{1}$$

Он характеризует долю вариации (разброса) зависимой переменной, объясненной с помощью данного уравнения. В качестве меры разброса зависимой переменной обычно используется ее дисперсия, а остаточная вариация может быть измерена как дисперсия отклонений вокруг линии регрессии. Если числитель и знаменатель вычитаемой из единицы дроби разделить на число наблюдений n, то получим, соответственно, выборочные оценки остаточной дисперсии и дисперсии зависимой переменной y. Отношение остаточной и общей дисперсий представляет собой долю необъясненной дисперсии. Если же эту долю вычесть из единицы, то получим долю дисперсии зависимой переменной, объясненной с помощью регрессии. Иногда при расчете коэффициента детерминации для получения несмещенных оценок дисперсии в числителе и знаменателе вычитаемой из единицы дроби делается поправка на число степеней свободы; тогда

$$R^2 = 1 - \left(\frac{\sum_i e_i^2}{n - m - 1} \right) : \left(\frac{\sum_i (y_i - \bar{y})^2}{n - 1} \right); \tag{2}$$

или, для парной регрессии, где число независимых переменных m равно 1,

$$R^2 = 1 - \left(\frac{\sum_i e_i^2}{n - 2} \right) : \left(\frac{\sum_i (y_i - \bar{y})^2}{n - 1} \right). \tag{3}$$

В числителе дроби, которая вычитается из единицы, стоит сумма квадратов отклонений наблюдений y_i от линии регрессии, в знаменателе - от среднего значения переменной y. Таким образом, дробь эта мала (а коэффициент R^2, очевидно, близок к единице), если разброс точек вокруг линии регрессии значительно меньше, чем вокруг среднего значения. МНК позволяет найти прямую, для которой сумма e_i^2 минимальна, а $y = \bar{y}$ представляет собой одну из возможных линий, для которых выполняется условие $\bar{y} = a + b\bar{x}$ ($b = 0$, $a = \bar{y}$). Поэтому величина в числителе вычитаемой из единицы дроби меньше, чем величина в ее знаменателе, - иначе выбираемой по МНК линией регрессии была бы прямая $y = \bar{y}$. Таким образом, коэффициент детерминации R^2 является мерой, позволяющей определить, в какой степени найденная регрессионная прямая дает лучший результат для объяснения поведения зависимой переменной y, чем просто горизонтальная прямая $y = \bar{y}$.

Смысл коэффициента детерминации может быть пояснен и немного иначе. Можно показать, что $\sum_i (y_i - \bar{y})^2 = \sum_i k_i^2 + \sum_i e_i^2$, где k_i - отклонение i-й точки на линии регрессии от $\bar{y}$. В данной формуле величина в левой части может интерпретироваться как мера общего разброса (вариации) переменной y, первое слагаемое в правой части $\sum_i k_i^2$ - как мера разброса, объясненного с помощью регрессии, и второе слагаемое $\sum_i e_i^2$ - как мера остаточного, необъясненного разброса (разброса точек вокруг линии регрессии). Если разделить эту формулу на ее левую часть и перегруппировать члены, то $R^2 = \dfrac{\sum_i k_i^2}{\sum_i (y_i - \bar{y})^2}$, то есть коэффициент детерминации R^2 есть доля объясненной части разброса зависимой переменной (или доля объясненной дисперсии, если разделить числитель и знаменатель на n или n-1).

Часто коэффициент детерминации R^2 иллюстрируют рис. 17.1.

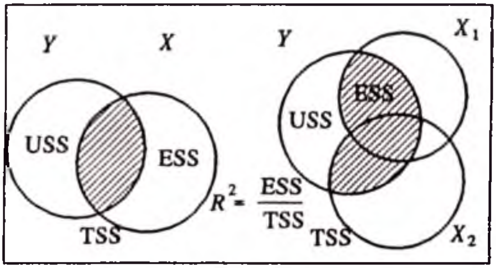


Рис. 17.1.

Здесь *TSS (Total Sum of Squares)* - общий разброс переменной y, *ESS (Explained Sum of Squares)* - разброс, объясненный с помощью регрессии, *USS (Unexplained Sum of Squares)* -разброс, необъясненный с помощью регрессии. Из рисунка видно, что с увеличением объясненной доли разброса коэффициент R^2 приближается к единице. Кроме того, из рисунка видно, что с добавлением еще одной переменной R^2 обычно увеличивается, однако если объясняющие переменные x_1 и x_2 сильно коррелируют между собой, то они объясняют одну и ту же часть разброса переменной y, и в этом случае трудно идентифицировать вклад каждой из переменных в объяснение поведения y.

Если существует статистически значимая линейная связь величин x и y, то коэффициент R^2 близок к единице. Однако он может быть близким к единице просто в силу того, что обе эти величины имеют выраженный временной тренд, не связанный с их причинно-следственной взаимозависимостью. В экономике обычно объемные показатели (доход, потребление, инвестиции) имеют такой тренд, а темповые и относительные (производительности, темпы роста, доли, отношения) - не всегда. Поэтому при оценивании линейных регрессий по временным рядам объемных показателей (например, зависимости выпуска от затрат ресурсов или объема потребления от величины дохода) величина R^2 обычно очень близка к единице. Это говорит о том, что зависимую переменную нельзя описать просто как равную своему среднему значению, но это и заранее очевидно, раз она имеет временной тренд.

Если имеются не временные ряды, а перекрестная выборка, то есть данные об однотипных объектах в один и тот же момент времени, то для оцененного по ним уравнения линейной регрессии величина R^2 не превышает обычно уровня 0,6-0,7. То же самое обычно имеет место и для регрессии по временным рядам, если они не имеют выраженного тренда. В макроэкономике примерами таких зависимостей являются связи относительных, удельных, темповых показателей: зависимость темпа инфляции от уровня безработицы, нормы накопления от величины процентной ставки, темпа прироста выпуска от темпов прироста затрат ресурсов. Таким образом, при построении макроэкономических моделей, особенно - по временным рядам данных, нужно учитывать, являются входящие в них переменные объемными или относительными, имеют ли они временной тренд.[1]

---

[1] В теории иногда встречаются модели, связывающие объемные и относительные показатели между собой. Например, это зависимость реальных инвестиций I от реальной ставки процента R: $I = a - bR$. Отметим, что эта зависимость может использоваться только в статической, краткосрочной модели. Если эту зависимость оценить по временным рядам, ничего хорошего обычно не получается. Показатель I в ней - объемный, и, следовательно, зависит от масштаба экономики в целом. Показатель R - относительный, и с масштабом экономики прямо не связан. Следовательно, если этого не учесть, то показатели инвестиций будут устойчиво отклоняться от линии регрессии на различных стадиях расчетного периода (направления отклонения зависят от динамики рассматриваемых переменных). Таким образом, в качестве объясняющего фактора нужно включить некоторый показатель, отражающий масштаб экономики (например, ВНП), либо просто добавить зависимость инвестиций от времени. Кроме масштаба экономики, важной для инвестирования является и предельная производительность капитала, которая меняется во времени. Таким образом, связь экономических переменных, которая адекватна для статической модели, далеко не всегда может быть оценена по рядам данных динамики.

Точную границу приемлемости показателя R^2 указать сразу для всех случаев невозможно. Нужно принимать во внимание и число степеней свободы уравнения, и наличие трендов переменных, и содержательную интерпретацию уравнения. Показатель R^2 может оказаться даже отрицательным. Как правило, это случается в уравнении без свободного члена $y = \sum_i a_i x_i$. Оценивание такого уравнения производится, как и в общем случае, по методу наименьших квадратов. Однако множество выбора при этом существенно сужается: рассматриваются не все возможные прямые или гиперплоскости, а только проходящие через начало координат. Величина R^2 получится отрицательной в том случае, если разброс значений зависимой переменной вокруг прямой (гиперплоскости) $y = \bar{y}$ меньше, чем вокруг даже наилучшей прямой (гиперплоскости) из проходящих через начало координат. Отрицательная величина R^2 в уравнении $y = \sum_i a_i x_i$ говорит о целесообразности введения в него свободного члена. Эта ситуация проиллюстрирована на рис. 17.2.

Линия 1 на нем - график уравнения регрессии без свободного члена (он проходит через начало координат), линия 2 - со свободным членом (он равен a_0), линия 3 - $y = \bar{y}$. Горизонтальная линия 3 дает гораздо меньшую сумму квадратов отклонений e_i, чем линия 1, и поэтому для последней коэффициент детерминации R^2 будет отрицательным.

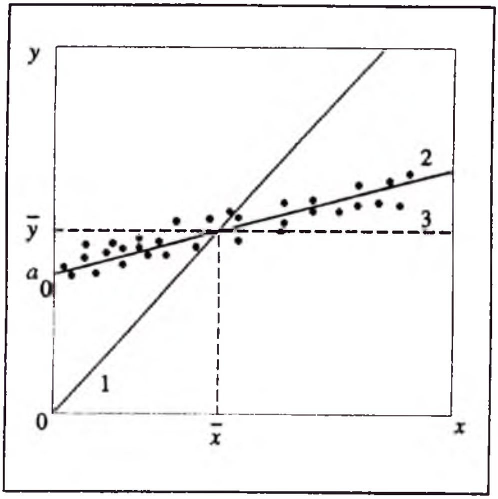


Рис. 17.2. Линии уравнений линейной регрессии $y=f(x)$ без свободного члена (1) и со свободным членом (2)

Поправка на число степеней свободы всегда уменьшает значение R^2, поскольку $(n-1)>(n-m-1)$. В результате величина R^2 также может стать отрицательной. Но это означает, что она была близкой к нулю до такой поправки, и объясненная с помощью уравнения регрессии доля дисперсии зависимой переменной очень мала.

17.2. *F*-статистика. Распределение Фишера в регрессионном анализе

Для определения статистической значимости коэффициента детерминации R^2 проверяется нулевая гипотеза для F-статистики, рассчитываемой по формуле:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} . \tag{4}$$

Соответственно, для парной регрессии $F = \frac{R^2(n - 2)}{1 - R^2}$. Смысл проверяемой гипотезы заключается в том, что все коэффициенты линейной регрессии, за исключением свободного члена, равны нулю. Если они действительно равны нулю для генеральной совокупности, то уравнение регрессии должно иметь вид $y = \bar{y}$, а коэффициент детерминации R^2 и F-статистика Фишера также равны нулю. При этом их оценки для случайной выборки, конечно, отличаются от нуля, но чем больше такое отличие, тем менее оно вероятно. Логика проверки нулевой гипотезы заключается в том, что если произошло событие, которое было бы слишком маловероятным в том случае, если данная гипотеза действительно была бы верна, то эта гипотеза отвергается.

Величина F, если предположить, что выполнены предпосылки относительно отклонений e_i, имеет распределение Фишера с $(m; n-m-1)$ степенями свободы, где m - число объясняющих переменных, n - число наблюдений. Распределение Фишера - двухпараметрическое распределение неотрицательной случайной величины, являющейся в частном случае, при $m=1$, квадратом случайной величины, распределенной по Стьюденту. Для распределения Фишера имеются таблицы критических значений, зависящих от чисел степеней свободы m и $n-m-1$, при различных уровнях значимости. В Приложении дана некоторая вспомогательная информация о распределении Фишера и показано, как пользоваться таблицами этого распределения.

Итак, показатели F и R^2 равны или не равны нулю одновременно, поэтому $F = 0$ равнозначно тому, что линия регрессии $y = \bar{y}$

является наилучшей по МНК и, следовательно, величина y статистически независима от x. Поэтому проверяется нулевая гипотеза для показателя F, который имеет хорошо известное, табулированное распределение - распределение Фишера. Для проверки этой гипотезы при заданном уровне значимости по таблицам находится критическое значение $F_{крит}$, и нулевая гипотеза отвергается, если $F > F_{крит}$. Пусть, например, при оценке парной регрессии по 15 наблюдениям $R^2 = 0{,}7$. В этом случае $F = 0{,}7 \cdot \frac{13}{0{,}3} \approx 30{,}3$. По таблицам для распределения Фишера с (1; 13) степенями свободы найдем, что при 5%-ном уровне значимости (доверительная вероятность 95%) критическое значение F равно 4,67, при 1%-ном - 9,07. Поскольку $F=30{,}3>F_{крит}$, нулевая гипотеза в обоих случаях отвергается. Если в той же ситуации $R^2 = 0{,}5$, то $F = 13$, и предположение о незначимости связи отвергается и здесь. Таким образом, для того, чтобы отвергнуть гипотезу о равенстве нулю одновременно всех коэффициентов линейной регрессии, коэффициент детерминации не должен быть очень близким к единице; его критическое значение для данного числа степеней свободы уменьшается при росте числа наблюдений и может стать сколь угодно малым. В то же время величина коэффициента R^2 (точнее, рассчитанной по нему F-статистики, поскольку последняя учитывает число наблюдений и число объясняющих переменных) может служить отражением общего качества регрессионной модели.

Отметим, что в случае парной регрессии проверка нулевой гипотезы для t - статистики коэффициента регрессии равносильна проверке нулевой гипотезы для F-статистики (и, соответственно, показателя R^2). В этом случае F-статистика равна квадрату t-статистики. В случае парной регрессии статистическая значимость величин R^2 и t-статистики коэффициента регрессии определяется коррелированностью переменных x и y. Самостоятельную важность показатель R^2 приобретает в случае множественной линейной регрессии.

Распределение Фишера может быть использовано не только для проверки гипотезы об одновременном равенстве нулю всех коэффициентов линейной регрессии, но и гипотезы о равенстве нулю части этих коэффициентов. Это особенно важно при развитии линейной регрессионной модели, так как позволяет оценить обоснованность исключения отдельных переменных или их групп из числа объясняющих переменных, или же, наоборот, включения их в это число.

Пусть, например, вначале была оценена множественная линейная регрессия $y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m$ по n наблюдениям с m объясняющими переменными, и коэффициент детерминации равен R_1^2. Затем последние k переменных исключены из числа объясняющих, и по тем же данным оценено уравнение $y = b_0 + b_1x_1 + b_2x_2 +$

... + $b_{m-k}x_{m-k}$, для которого коэффициент детерминации равен R_2^2 (он обязательно уменьшился, поскольку каждая дополнительная переменная объясняет часть, пусть небольшую, вариации зависимой переменной). Для того чтобы проверить гипотезу об одновременном равенстве нулю всех коэффициентов регрессии при исключенных переменных, рассчитывается величина $F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - m - 1}{k}$, имеющая распределение Фишера с (k, n-m-1) степенями свободы. По таблицам, при заданном уровне значимости, находится критическое значение F-статистики, и если ее рассчитанное значение превосходит критическое, то нулевая гипотеза отвергается. В таком случае исключать сразу из числа объясняющих все k переменных некорректно. F-статистика оказывается относительно большой, если велика разность (R_1^2-R_2^2). В этом случае исключение данного набора k объясняющих переменных приводит к слишком большому сокращению доли объясненной дисперсии зависимой переменной, и поэтому недопустимо. Если, наоборот, эта доля сокращается незначительно, то F-статистика невелика, нулевая гипотеза не отвергается, и указанные k переменных могут быть исключены из уравнения регрессии. Аналогичные рассуждения могут быть проведены и по поводу обоснованности включения в уравнение регрессии одной или нескольких (k) новых объясняющих переменных. В этом случае рассчитывается F-статистика $F = \frac{R_2^2 - R_1^2}{1 - R_2^2} \cdot \frac{n - m - k - 1}{k}$, имеющая распределение $F(k, n$-m-k-1), и если она превышает критический уровень, то включение новых переменных объясняет существенную часть необъясненной ранее дисперсии зависимой переменной y. Отметим лишь, что добавлять новые переменные целесообразно, как правило, по одной.

В вопросе о добавлении объясняющих переменных в уравнение регрессии полезным может оказаться рассмотрение R^2 с поправкой на число степеней свободы: $R^2 = 1 - \left(\frac{\sum_i e_i^2}{n - m - 1} \right) : \left(\frac{\sum_i (y_i - y)^2}{n - 1} \right)$.

Обычный R^2 (без поправки) всегда растет при добавлении новой переменной; в R^2 с поправкой растет величина m, уменьшающая его. Если увеличение доли объясненной дисперсии при добавлении новой переменной мало, то R^2 с поправкой может уменьшиться. Если это так, то добавлять переменную нецелесообразно.

F-статистика Фишера используется также для проверки гипотезы о совпадении уравнений регрессии для отдельных групп наблюдений. Пусть имеются две выборки, содержащие, соответственно, n_1 и n_2 наблюдений. Для каждой из этих выборок оценено уравнение регрессии вида $y = a_0 + a_1x_1 + a_2x_2 + ... + a_mx_m$. Пусть суммы квадратов отклонений y_i от линий регрессии равны для них, соответственно, S_1 и S_2. Проверяется нулевая гипотеза, заключающаяся в том, что все соответствующие коэффициенты этих уравнений равны друг другу, то есть что уравнение регрессии для этих выборок одно и то же. Пусть оценено уравнение регрессии того же вида сразу для всех (n_1+n_2) наблюдений, и сумма квадратов отклонений y_i от линии регрессии равна для него S_0. Тогда рассчитывается F-статистика по формуле $F = \frac{(S_0 - S_1 - S_2)}{(S_1 + S_2)} \cdot \frac{(n_1 + n_2 - 2m - 2)}{(m + 1)}$. Она имеет распределение Фишера с $(m+1, n_1+n_2-2m-2)$ степенями свободы. F-статистика будет близкой к нулю, если уравнение регрессии для обеих выборок одинаково, поскольку в этом случае $S_0=S_1+S_2$. Если же ее расчетное значение велико (то есть больше критического значения при данном уровне значимости), то нулевая гипотеза отвергается. Описанная процедура важна для ответа на вопрос, можно ли за весь рассматриваемый в модели период времени построить единое уравнение регрессии, или же нужно разбить его на части и на каждой из частей строить свое уравнение регрессии.

17.3. Проверка условий, выполнение которых предполагалось при оценивании уравнения регрессии. Автокорреляция остатков. Статистика Дарбина-Уотсона

Близкое к единице значение коэффициента детерминации R^2 еще не свидетельство высокого качества уравнения регрессии. Рассмотрим рис. 17.3. На нем показана зависимость реального объема потребления (*CONS*, млрд.долл., 1982 г.) от численности населения (*POP*, млн.) в США за 1931-1990 гг., а также линия оцененного по этим данным уравнения парной линейной регрессии. Формула этого уравнения следующая:

$$CONS = -1817{,}3 + 16{,}7\ POP. \tag{5}$$

Стандартные ошибки свободного члена и коэффициента регрессии равны, соответственно, 84,7 и 0,46; их t-статистики - (-21,4 и 36,8). По абсолютной величине t-статистики намного превышают 3, и это свидетельствует о высокой надежности оцененных коэффициентов. Коэффициент детерминации R^2 уравнения равен 0,96, то есть объяснено 96% дисперсии объема потребления. И в то же время уже по рисунку видно, что оцененная регрессия не очень хоро-

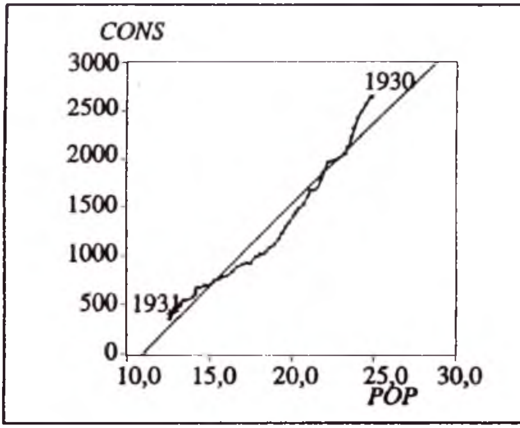


Рисунок 17.3. График зависимости реального объема потребления (*CONS*, млрд. долл., 1982 г.) от численности населения (*POP*, млн.) в США в 1931-1990 гг.

ша: зависимость величин *POP* и *CONS* явно нелинейна. Если использовать проведенную прямую, скажем, для прогнозирования дальнейшей динамики потребления, результат будет неудовлетворительным. Существо вопроса здесь понятно - в течение рассматриваемого периода значительно вырос объем потребления в расчете на душу населения. Численность населения США росла во времени почти линейно (то есть с постоянными годовыми приростами), а объем потребления - по экспоненте (то есть с примерно постоянным темпом). Это ясно и без уравнения линейной регрессии, но мы специально оценили его для иллюстрации.

Как же можно выразить формально неудовлетворительность полученного уравнения регрессии? Можно видеть, что не выполнены необходимые предпосылки об отклонениях от линии регрессии e_i. Эти величины явно не являются взаимно независимыми, и дисперсия их не постоянна. Нарушения исходных предпосылок не только свидетельствуют о неточной спецификации уравнения регрессии, но и делают неточными полученные оценки коэффициентов регрессии и их стандартных ошибок. Поэтому следующий этап проверки качества уравнения регрессии - проверка некоторых важных свойств, выполнение которых предполагалось при оценивании уравнения регрессии.

Приступая к оценке линейного уравнения регрессии, мы предполагали, что реальная взаимосвязь переменных линейна, а отклонения от регрессионной прямой случайны, независимы между собой и имеют нулевое среднее и постоянную дисперсию. Так ли это на самом деле? Если нет, то наш анализ статистической значимости коэффициентов регрессии неточен и оценки этих коэффициентов не обладают такими желательными свойствами, как несмещенность, состоятельность и эффективность.

Попытаемся ответить на вопрос, в каких случаях отклонения не обладают предполагавшимися свойствами. Во-первых, если в действительности исследуемая взаимосвязь нелинейна. Мы видим, например, на рис. 17.3, что в этом случае отклонения от линии регрессии не случайно распределены вокруг нее, а обладают определенной закономерностью. Эта закономерность, в частности, выражается в одинаковом, как правило, знаке каждых двух соседних отклонений. Это может являться следствием нелинейного характера связи переменных, либо воздействием какого-то фактора, не включенного в уравнение регрессии. Величина такого неучтенного фактора может менять свою динамику в рассматриваемый период, отклоняясь в достаточно длительные промежутки времени в ту или иную сторону от своего среднего значения. Это, очевидно, может служить причиной длительных устойчивых отклонений зависимой переменной от линии регрессии. Обе указанные причины свидетельствуют о том, что существует возможность улучшить уравнение регрессии путем оценивания какой-то новой нелинейной формулы или включения некоторой новой объясняющей переменной.

Зависимость, показанная на рис. 17.3, очевидно, нелинейна. Но это - крайний случай. Далеко не всегда бывает столь же очевидно, что отклонения от регрессионной прямой имеют неслучайный, закономерный характер. Для оценки степени такой неслучайности необходимо ввести количественную меру.

Итак, одним из основных предполагаемых свойств отклонений e_i значений y_i от регрессионной формулы $y=\alpha+\beta x$ является их статистическая независимость между собой. Поскольку значения e_i остаются неизвестными ввиду неизвестности истинных значений б и в, то проверяется статистическая независимость их аналогов - отклонений e_i. При этом проверяется обычно их некоррелированность (являющаяся необходимым, но недостаточным атрибутом независимости), причем некоррелированность не любых, а соседних величин e_i. Соседними можно считать соседние во времени (в случае временных рядов) или по возрастанию переменной x (в случае перекрестных выборок) значения e_i. Для этих величин можно рассчитать, например, коэффициент корреляции (называемый *коэффициентом автокорреляции первого порядка*):

$r_{i,i-1} = \dfrac{\sum_i e_i e_{i-1}}{\sqrt{\sum_i e_i^2 \sum_i e_{i-1}^2}}$ (считаем, что $M[e_i] = 0$). Практически, однако, используют тесно связанную с $r_{i,i-1}$ статистику Дарбина-Уотсона DW, рассчитываемую по формуле:

$$DW = \frac{\sum_i (e_i - e_{i-1})^2}{\sum_i e_i^2}. \quad (6)$$

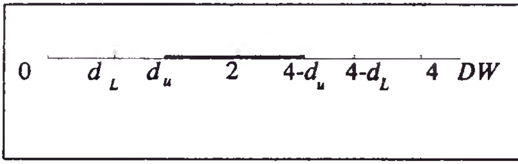
Очевидно, $DW = \dfrac{\sum_i e_i^2 - 2\sum_i e_i e_{i-1} + \sum_i e_{i-1}^2}{\sum_i e_i^2}$, и, поскольку при больших n $\sum_i e_i^2 \approx \sum_i e_{i-1}^2$, получаем $DW \approx 2(1-r_{i,i-1})$. Если e_i в точности равно e_{i-1}, то $DW = 0$; если $e_i = -e_{i-1}$, то $DW = 4$, во всех других случаях $0 < DW < 4$.

В случае, когда каждое отклонение e_i примерно совпадает с предыдущим отклонением e_{i-1}, каждое слагаемое в числителе величины DW близко к нулю. Сумма квадратов разностей отклонений в числителе будет намного меньше суммы квадратов отклонений в знаменателе, и поэтому статистика Дарбина-Уотсона окажется близкой к нулю. Рис. 17.3 представляет такой случай - это случай положительной автокорреляции остатков первого порядка. Значение статистики Дарбина-Уотсона здесь равно 0,045, что очень мало и подтверждает статистическую зависимость отклонений e_i без всяких таблиц. Другой крайний случай возникает, когда точки наблюдений поочередно отклоняются в разные стороны от линии регрессии, и каждое следующее отклонение e_i имеет, как правило, противоположный знак, чем предыдущее отклонение e_{i-1}. В этом случае $(e_i - e_{i-1}) \approx 2e_i$, и $DW \approx \dfrac{\sum_i (2e_i)^2}{\sum_i e_i^2} = 4\dfrac{\sum_i e_i^2}{\sum_i e_i^2} = 4$. Это - случай отрицательной автокорреляции остатков первого порядка. Последняя достаточно редко встречается в экономическом анализе. Если рассматриваются временные ряды с годовыми данными, то подобную закономерность поведения последовательных отклонений довольно трудно проинтерпретировать. Однако она может встретиться при работе, например, с полугодовыми данными показателей с сезонным характером изменения. Наконец, если характер поведения отклонений случаен, можно предположить, что в половине случаев знак последовательных отклонений совпадает, а в половине - различен. Поскольку абсолютная величина их в среднем предполагается одинаковой, можно считать, что здесь в половине случаев e_i равно e_{i-1}, а в оставшейся половине e_i равно $-e_{i-1}$.

Итак, при этом $DW \approx \dfrac{\sum_i 0{,}5\cdot(2e_i)^2}{\sum_i e_i^2} = 0{,}5\cdot 4\cdot\dfrac{\sum_i e_i^2}{\sum_i e_i^2} = 2$. Это показывает, что близость статистики Дарбина-Уотсона к двум является необходимым условием случайного характера отклонений от линии регрессии. Нужно, однако, иметь в виду, что в показателе *DW* сравниваются только соседние отклонения от регрессии, в то же время циклы изменения экономических переменных могут быть более или менее длительными, чем одна единица времени. Например, если рассматриваются поквартальные данные сельскохозяйственного производства (имеющего годовой цикл) и оценивается их линейная регрессия от времени, статистика Дарбина-Уотсона может быть близкой к двум при выраженной регулярности отклонений зависимой переменной от линии регрессии.

Если статистика Дарбина-Уотсона близка к двум, мы считаем отклонения от регрессии случайными (хотя в действительности они могут и не быть таковыми). Это означает, что линейная функция, вероятно, отражает реальную взаимосвязь; скорее всего, не осталось существенных неучтенных факторов, влияющих на зависимую переменную, и какая-либо другая, нелинейная формула не превосходит по статистическим характеристикам данную линейную. Даже если доля дисперсии зависимой переменной, объясненной с помощью регрессии, при этом мала, можно ожидать, что другая часть этой дисперсии, оставшаяся необъясненной, порождена действием множества различных малых факторов и может быть описана как случайная нормальная ошибка. Но как определить, достаточно ли близка величина статистики *DW* к двум? Для этого имеются специальные таблицы, позволяющие при данном числе наблюдений и объясняющих переменных, для заданного уровня значимости, найти критические значения статистики Дарбина-Уотсона.

Итак, статистика Дарбина-Уотсона применяется для проверки гипотезы об отсутствии автокорреляции остатков e_i первого порядка (нулевой гипотезы). Для этого по таблицам находятся (при данном уровне значимости, числе наблюдений и независимых переменных) доверительные интервалы, в пределах которых нулевая гипотеза принимается, отвергается или не может быть принята или отвергнута. Важно, что для статистики Дарбина-Уотсона существуют два критических значения, меньшие двух: нижнее d_l как граница для признания положительной автокорреляции остатков и верхнее d_u как граница признания ее отсутствия. Для проверки гипотезы об отрицательной автокорреляции остатков эти критические значения отражаются симметрично относительно числа 2:



Например, пусть оценена парная линейная регрессия по 15 наблюдениям, и $DW = 1,1$. Зададим уровень значимости 5% и найдем по таблицам $d_l = 0,95$;

$d_u = 1,23$. Нулевая гипотеза была бы принята при $d_u = 1,23 < DW < 2,77 = 4\text{-}d_u$ и отвергнута при $DW < 0,95 = d_l$ или $DW > 3,05 = 4 - d_l$. Поскольку в данном случае DW лежит между d_u и d_l, нулевая гипотеза не может быть ни принята, ни отвергнута. Если альтернативной гипотезой является гипотеза о положительной автокорреляции остатков (отрицательная из содержательных соображений отбрасывается), то критические значения $d_u = 1,23$ и $d_l = 0,95$ соответствуют 2,5%-ному уровню значимости.

Как в общем случае выглядят примерно критические величины статистики DW? Очень грубо, в первом приближении можно сказать, что при достаточном числе наблюдений (не меньше 12-15), при 1-3 объясняющих переменных DW должна быть не менее 1 (и не больше 3). В противном случае мы признаем существование автокорреляции остатков и попытаемся улучшить формулу. Если статистика DW находится приблизительно между 1,2-1,3 и 2,7-2,8, мы можем считать, что статистически значимая автокорреляция остатков отсутствует. В промежуточном случае достаточно надежный вывод сделан быть не может. Если число наблюдений растет, то критические значения статистики Дарбина-Уотсона d_l и d_u приближаются к двум: для 60-70 наблюдений ее нижнее критическое значение d_l составляет примерно 1,4-1,5. Это верно для прежнего относительно малого числа объясняющих переменных; если это число растет, то критическое значение DW становится меньше.

Итак, обобщая, если статистика Дарбина-Уотсона составляет 1,5-2,0-2,5, мы хотя и не можем быть абсолютно уверены, что отклонения от линии регрессии взаимно независимы, но обычно удовлетворяемся этим в проверке их независимости.

В случае наличия автокорреляции остатков полученная формула регрессии считается обычно неудовлетворительной. Взглянув на график поведения отклонений e_i, можно поискать другую (нелинейную) формулу, включить неучтенные до этого факторы, уточнить период проведения расчетов или разбить его на части, либо применить к данным уменьшающее автокорреляцию остатков преобразование (например, автокорреляционное преобразование или метод скользящих средних). Так, в рассмотренной уже зависимости объема реального потребления от численности населения объясняющая переменная *POP* должна быть заменена другой; обычно это объем располагаемого дохода Y_d. Если добавить переменную Y_d к оцененному уравнению, переменная *POP* становится незначимой (ее t-статистика равна 0,16). Высокий уровень R^2 в первоначальном уравнении был обусловлен не тем, что динамика численности населения определяла динамику объема реального потребления, а тем, что обе эти переменные имели выраженную тенденцию возрастания в рассматриваемый период.

Статистика *DW* позволяет проверить некоррелированность отклонений от линии регрессии. Некоторые другие свойства этих отклонений (например, постоянство их дисперсии) могут быть также проверены с помощью специальных статистик. Мы не будем останавливаться на этом подробно, упомянув лишь о существовании самой проблемы. Рассуждения при этом могут быть подобными прежним: если значения тестовых статистик "плохие", то можно попытаться уточнить формулу связи, набор объясняющих переменных или процедуру оценивания.

17.4. Прогнозирование

Рассмотрим зависимость объема реального частного потребления в США (*CONS*) от располагаемого дохода (*DINC*) за 1971-1990 гг., приведенную в предыдущей главе (см. рис. 16.6). Эта зависимость имеет вид

$$\underset{}{CONS} = \underset{(-7,7)}{-217,6} + \underset{(81,9)}{1,007}\cdot DINC \qquad (7)$$

(в скобках приведены t-статистики)
R^2=0,997; *DW*=1,58.

Свободный член и коэффициент регрессии здесь статистически значимы, а коэффициент детерминации R^2 очень высок; для анализа качества этих показателей нет нужды прибегать к таблицам. По таблицам статистики *DW* найдем, что при 20 наблюдениях, одной объясняющей переменной и 5%-ном уровне значимости d_l=1,08; d_u=1,28. Поскольку 2>*DW*=1,58>1,28, гипотеза об отсутствии автокорреляции остатков первого порядка не отвергается. Таким образом, со статистической точки зрения данная зависимость приемлема по всем показателям. На рис. 17.4 приведены для нее графики фактических значений зависимой переменной (сплошная линия в верхней части рисунка), оцененных по уравнению регрессии ее значений (пунктирная линия) и отклонений e_i (нижняя часть рисунка).

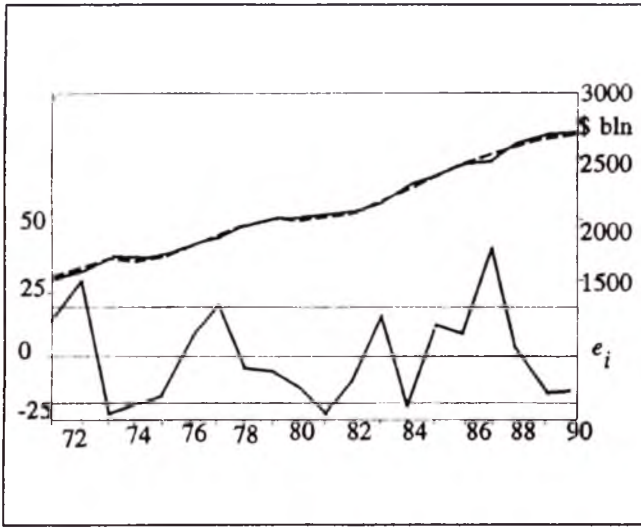


Рисунок 17.4. Объем реального частного потребления в США (*CONS*, млрд.долл., 1982 г.) в 1971-1990 гг; фактический и оцененный по уравнению регрессии. Отклонения от линии регрессии

На рис. 17.4 можно видеть, что рассчитанные по уравнению регрессии значения переменной *CONS* довольно близки к фактическим ее значениям. Стандартная ошибка регрессии $S = \sqrt{\frac{\sum_i e_i^2}{n-2}}$ примерно равна 20 при среднем значении зависимой переменной около 2000, то есть составляет около 1%. Отклонения от линии регрессии носят случайный характер, и их среднее значение остается приблизительно постоянным.

Отношение стандартной ошибки регрессии к среднему значению зависимой переменной $V = \frac{S}{\bar{y}}$ может служить критерием прогнозных качеств оцененной регрессионной модели. Если величина V мала и отсутствует автокорреляция остатков (то есть систематичность отклонений зависимой переменной от линии регрессии), проверяемая с помощью статистики DW, то прогнозные качества модели высоки. Если уравнение регрессии используется в прогнозировании, то величина V часто рассчитывается не для того периода, на котором было оценено уравнение, а для некоторого следующего за ним, "постпрогнозного", периода, для которого имеются наблюдения зависимой и объясняющих переменных. Тем самым прогнозные качества модели проверяются на практике. И уже для последующего периода, если для него известны прогнозы значений объясняющих переменных, может быть построен прогноз зависимой переменной. Считается, что период прогнозирования должен быть по крайней мере в 3 раза короче, чем тот период, для которого было оценено уравнение регрессии.

Для примера оценим функцию зависимости *CONS* от *DINC* за период не 1971-1990, а 1971-1986 гг., а затем построим постпрогноз (то есть прогноз, делаемый "задним числом") на период 1987-1990 гг. Уравнение регрессии получается следующее, приемлемое по всем параметрам:

$$CONS = \underset{(-5{,}6)}{-208{,}8} + \underset{(58.8)}{1{,}003} \cdot Y_d \qquad (8)$$

(в скобках приведены t-статистики)
R^2=0,996; DW=1,72.

График "постпрогнозных" значений объема потребления *CONSF* показан, наряду с графиком его фактических значений *CONS*, на рис. 17.5.

В целом прогноз оказался довольно удачным; лишь в 1987 году его ошибка довольно велика и составляет примерно 5%.

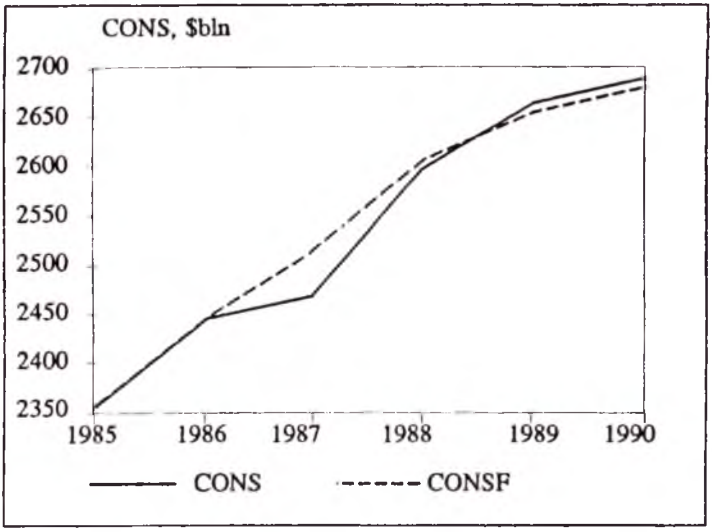


Рисунок 17.5. Постпрогноз реального частного потребления в США в 1987-1990 гг. (*CONSF*) и его фактическая динамика (*CONS*)

Оценим прогнозные качества модели более точно, рассчитав среднюю относительную ошибку прогноза V. Поскольку для постпрогнозного периода число степеней свободы равно числу точек k=4, стандартная ошибка прогноза на 1987-1990 гг. рассчитывается как $S = \sqrt{\frac{\sum_i e_i^2}{4}} = 25{,}2$. Относительная ошибка прогноза $V = \frac{S}{\bar{y}} = \frac{25{,}2}{2615{,}3}$ $= 0{,}96\%$. Если относительную ошибку прогноза оценить по расчетному периоду 1971-1986 гг., то она оказывается равной $V = \frac{S}{\bar{y}} =$ $\frac{17{,}5}{1947{,}5} = 0{,}90\%$, где $S = \sqrt{\frac{\sum_i e_i^2}{16-2}} = 17{,}5$. Таким образом, оценка прогнозных качеств уравнения регрессии дает хороший, примерно одинаковый результат (менее 1% ошибки) как на расчетном, так и на контрольном (постпрогнозном) периоде. Для построения прогноза объема потребления на период после 1990 года нужно оценить уравнение регрессии за 1971-1990 гг. (что уже сделано выше) и подставить в него прогнозируемые значения величины располагаемого дохода.

17.5. Модель инфляции. Эконометрическая оценка NAIRU

В заключение главы рассмотрим еще один пример использования модели множественной линейной регрессии в макроэкономическом анализе. Предположим, что у нас есть ряды статистических данных уровней инфляции и безработицы в экономике США за 1977-1990 гг., и что мы хотим оценить взаимосвязь между этими показателями. Из теоретических соображений мы можем предполагать, что темп инфляции отрицательно связан с уровнем безработицы и, кроме того, обладает определенной инерцией. Мы также предполагаем, что существует некоторый “естественный” уровень безработицы, не воздействующий на темп инфляции. Этот уровень называется *NAIRU - Not Accelerating Inflation Rate of Unemployment (“уровень безработицы, не ускоряющий инфляцию”)*. Предположим в соответствии со сказанным, что формула зависимости этих показателей имеет вид

$$INF = a\,(u^* - u) + b\,INF(-1), \tag{9}$$

где INF - темп инфляции, $INF(-1)$ - темп инфляции предыдущего года (инерционная составляющая, описывающая ожидаемую инфляцию), u - уровень безработицы, u^* - NAIRU, a и b - неизвестные коэффициенты. Для указанных временных рядов было оценено следующее уравнение множественной линейной регрессии:

$$\underset{}{INF} = \underset{(0{,}930)}{5{,}414} - \underset{(0{,}154)}{0{,}920u} + \underset{(0{,}086)}{1{,}148 INF(-1)} \tag{10}$$

$R^2 = 0{,}942$; $DW = 2{,}07$.

В скобках указаны стандартные ошибки соответствующих коэффициентов. Можно отметить, что статистическое качество полученного уравнения регрессии практически идеально. Все t-статистики превышают 5 по абсолютной величине (а, грубо говоря, границей для очень хорошей оценки является 3). Очень высока доля дисперсии зависимой переменной, объясненная с помощью уравнения регрессии, - 94,2% - особенно с учетом того, что уравнение регрессии связывает относительные величины, не имеющие выраженного временного тренда. Статистика Дарбина-Уотсона DW очень близка к 2, и, даже не прибегая к таблицам, здесь ясно, что гипотеза об отсутствии автокорреляции остатков первого порядка будет принята при любом разумно малом уровне значимости. Итак, мы имеем хороший пример линейной регрессии, когда можно оценить ее статистическую значимость, не прибегая к таблицам распределений Стьюдента, Фишера или Дарбина-Уотсона, а лишь по общему порядку полученных статистик.

Теперь рассмотрим содержательный смысл и теоретическую интерпретацию полученного уравнения. Оно показывает, что инфляция является очень инерционным процессом, и то, что коэффициент b=1,148 превышает единицу, говорит о том, что это процесс самоускоряющийся. Последнее существенно усложняет задачу контролирования и сдерживания инфляции. Учитывая,что в соответствии с нашей моделью, $5{,}414 - 0{,}920u = a(u^* - u)$, получаем $a = 0{,}920$ и $u^* = 5{,}414/0{,}920 = 5{,}885$ (%). Итак, если мы принимаем исходную общую формулу модели инфляции, то мы тем самым получаем из нее оценку естественного уровня безработицы (*NAIRU*) для экономики США в 1977 - 1990 гг. Она оказалась несколько меньшей, чем 6%, что согласуется с другими оценками, полученными различными методами.

ПРИЛОЖЕНИЕ
F-распределение Фишера

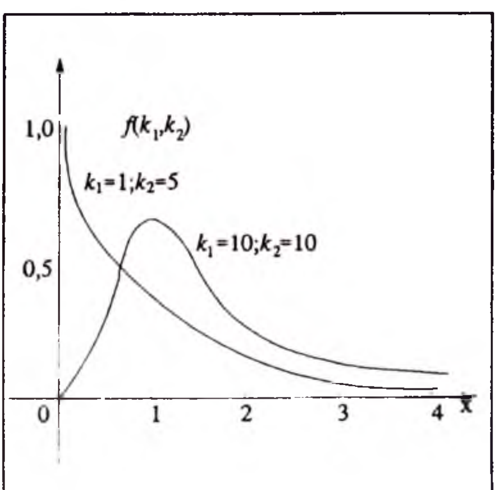
Это распределение (называемое иногда распределением дисперсионного отношения) имеет случайная величина, равная отношению двух независимых случайных величин: величины $\frac{\chi^2(k_1)}{k_1}$, выражающейся через случайную величину, имеющую распределение χ^2 с k_1 степенями свободы и величины $\frac{\chi^2(k_2)}{k_2}$, выражающейся через случайную величину, имеющую распределение χ^2 с k_2 степенями свободы.

Вводя новую случайную величину $F(k_1,k_2) = \frac{\chi^2(k_1)}{k_1} : \frac{\chi^2(k_2)}{k_2}$, мы получим для нее распределение Фишера с k_1 и k_2 степенями свободы с плотностью вероятности:

$$f_F(x,k_1,k_2) = Cx^{\frac{k_1}{2}-1}\left(1 + \frac{k_1}{k_2}x\right)^{-\frac{k_1+k_2}{2}},$$

$$M|X| = \frac{k_2}{k_2 - 2}\ (k_2>2),\ D(X)=h(k_1,k_2).$$

При больших k_1 и k_2 это распределение приближается к нормальному.



Критические точки распределения Фишера обладают следующим свойством:

$$F_{1-\alpha}(k_1,k_2) = \frac{1}{F_\alpha(k_2,k_1)} .$$

Квадрат случайной величины, имеющей распределение Стьюдента с k_2 степенями свободы имеет распределение Фишера с $(1,k_2)$ степенями свободы.

Подставляя в определение случайной величины F "выборочное представление" случайной величины χ^2: $\chi^2(n-1) = \dfrac{(n-1)S_n^{**2}}{\sigma^2}$, можно получить "выборочное представление" случайной величины F:

$$F(k_1-1,\ k_2-1) = \left(\frac{S_{1,k_1}^{*2}}{\sigma_1^2}\right) : \left(\frac{S_{2,k_2}^{*2}}{\sigma_2^2}\right),$$

где $S_n^{**2} = \dfrac{n}{n-1}S_n^{*2} = \dfrac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$ - исправленная выборочная дисперсия для выборки объема n.

Распределение Фишера используется, например, при:

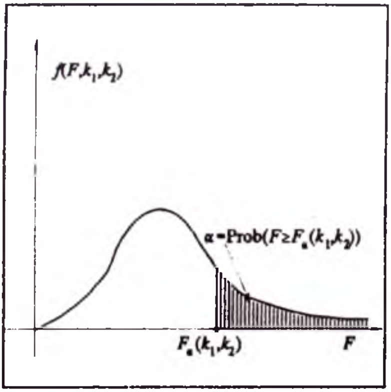
- сравнении двух дисперсий;
- проверке гипотезы об одновременном равенстве нулю всех или части коэффициентов линейной регрессии;
- проверке гипотезы о совпадении всех коэффициентов двух уравнений линейной регрессии.

Работа с таблицами F-распределения Фишера

Таблицы функции F-распределения Фишера на интервале $[0,+\infty)$ обычно приводятся отдельно для различных значений вероятности α попадания в "хвост" функции распределения. Например, для $\alpha = 0{,}05$ такая таблица имеет вид

k2 \ k1	1	...	10	...	100	∞
1	161	...	242	...	253	254
...	...	...	...	...	...	...
10	4,96	...	2,97	...	2,59	2,54
...	...	...	...	...	...	...
100	3,94	...	1,92	...	1,39	1,28
∞	3,84	...	1,83	...	1,24	1,00

В этой таблице для различных сочетаний чисел степеней свободы k_1 и k_2 приведены критические точки функции распределения Фишера, соответствующие вероятности $\alpha = 0{,}05$ попадания в "хвост" функции распределения.

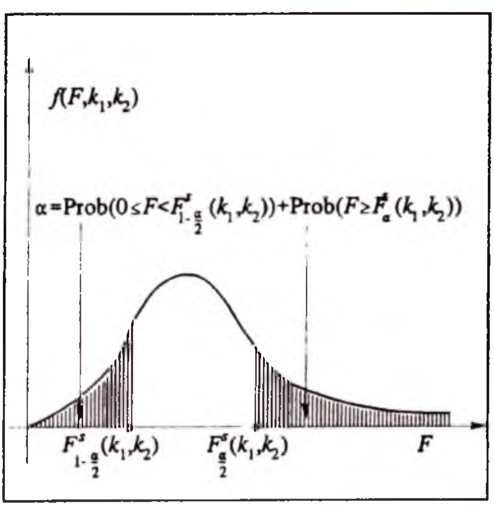


Критическая точка $F_\alpha(k_1,k_2)$, например, $F_{0{,}05}(10,\ 100)$, находится в таблице, соответствующей значению $\alpha = 0{,}05$, на пересечении строки k_2 (в данном случае $k_2 = 100$) и столбца k_1 (в данном случае $k_1 = 10$). Из приведенной таблицы находим, что $F_{0{,}05}(10,\ 100) = 1{,}92$. Напомним, что критическая точка в данном случае имеет следующий смысл: $\mathrm{Prob}\{F > F_\alpha(k_1,k_2)\} = \alpha$.

Отметим, что иногда таблицы F-распределения приводятся для двусторонних критических точек $F_\alpha^s(k_1,k_2)$, определяемых из условия

$$\mathrm{Prob}\{F_{1-\alpha/2}^s(k_1,k_2)\ F < F_{\alpha/2}^s(k_1,k_2)\} = 1 - \alpha.$$

Появление здесь величины $\alpha/2$ объясняется тем, что при заданной вероятности α попадания в оба "хвоста" функции распределения вероятность попадания в каждый из "хвостов" функции распределения обычно считается одинаковой. Следовательно, она в два раза меньше α и равна $\alpha/2$.



Вопросы к главе 17

1. Из каких этапов состоит проверка качества оцененного уравнения регрессии?
2. Как рассчитывается и что показывает коэффициент детерминации R^2?
3. Коэффициент детерминации $R^2 = 0{,}5$. Что можно сказать о качестве оцененной формулы в целом? Какая нужна дополнительная информация?
4. Что такое распределение Фишера? В каких задачах эконометрики оно используется?
5. Если нулевая гипотеза для статистики Фишера отвергается, то что можно сказать про оцененную парную линейную регрессию?
6. Таблицы каких распределений используются при оценке качества линейной регрессии?
7. Какие показатели характеризуют независимость отклонений зависимой переменной от линии регрессии? Как осуществляется проверка этой независимости?
8. В каких случаях наблюдается положительная автокорреляция отклонений e? Приведите примеры из экономики.
9. Статистика Дарбина-Уотсона оказалась близкой к четырем. Что это означает?
10. Если y зависит от x как квадратичная функция $y = x^2$, но оценена связывающая их линейная регрессия, то какой окажется величина DW?
11. Как осуществляется прогнозирование экономических показателей с использованием моделей линейной регрессии?
12. Как можно оценить “естественный” уровень безработицы с использованием модели линейной регрессии?

ГЛАВА 18

ПОСТРОЕНИЕ И РАЗВИТИЕ МОДЕЛИ ЛИНЕЙНОЙ РЕГРЕССИИ. ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ МАКРОЭКОНОМИЧЕСКИХ МОДЕЛЕЙ

18.1. Направления совершенствования линейной регрессионной модели

Модель чистого экспорта

Как уже отмечалось в предыдущих главах, построение и развитие эконометрической модели - это длительный и сложный процесс. Очень редко оценка исходной спецификации зависимости дает хорошие по всем параметрам результаты. Предположим, что оцененная множественная линейная регрессия по ряду статистических характеристик (*DW*, *t*-статистики, *F*-статистика) оказалась неприемлемой и требует уточнения. Направления такого уточнения могут быть следующими:

- выведение из рассмотрения незначимых объясняющих переменных и добавление новых переменных;
- разбиение временного интервала на части и оценка исходной или новой формулы регрессии на каждой из них;
- преобразование исходных данных с целью устранить их нежелательные свойства;
- построение нелинейных спецификаций уравнения регрессии с последующей их линеаризацией (или оценкой нелинейной регрессии);
- устранение сильно коррелированных между собой объясняющих переменных (борьба с мультиколлинеарностью).

Мы рассмотрим эти направления совершенствования регрессионной модели на примере конкретного эконометрического исследования, делая по мере необходимости пояснения и отступления. В качестве базового примера рассмотрим процесс построения функции чистого экспорта для экономики США. Для этого будем использовать массив макроэкономических данных за 1931 - 1990 гг. В качестве первоначальной спецификации функции чистого экспорта для этого периода в целом рассмотрим выражение

$$RNX = c + b_1 \cdot GNP + b_2 \cdot RSR. \qquad (1)$$

Здесь переменная *RNX* обозначает реальный чистый экспорт (*Real Net Exports*), или чистый экспорт в постоянных ценах 1982 г., млрд. долларов; *GNP* - реальный валовой национальный продукт в тех же единицах; *RSR* - реальная краткосрочная процентная ставка, в процентах. В различные макромодели открытой экономики, в частности в модель *IS-LM*, обычно включаются зависимости чистого экспорта такого или подобного вида. Коэффициенты b_1 и b_2, называемые чувствительностями величины чистого экспорта к показателю объема ВНП и величине ставки процента, считаются в теории отрицательными. В соответствии с результатами оценивания на каждом очередном шаге мы будем корректировать совокупность объясняющих переменных, период оценивания и другие особенности уравнения (временные лаги, наличие свободного члена и т.д.).

Оценка первоначальной формулы дает результат

$$RNX = \underset{(8{,}43)}{21{,}1} - \underset{(0{,}004)}{0{,}017} \cdot GNP - \underset{(0{,}947)}{0{,}411} \cdot RSR \qquad (2)$$

(в скобках приведены стандартные ошибки)
$R^2 = 0{,}29$; $DW = 0{,}43$.

Отрицательные знаки коэффициентов регрессии соответствуют здесь теоретическим представлениям. Коэффициент при переменной *GNP* значительно меньше по абсолютной величине, чем коэффициент при *RSR*, но это не значит, что данная величина воздействует на зависимую переменную слабее. Здесь все определяется единицами измерения, и если ВНП измерять не в миллиардах, а в триллионах долларов, то соответствующий коэффициент регрессии будет равен не 0,017, а 17, при стандартной ошибке 4.

Соотношение коэффициента и его стандартной ошибки, или t-статистика (в последнем случае 0,017:0,004 = 4,25), важна для определения статистической значимости зависимости функции от соответствующей объясняющей переменной. Вообще говоря, нулевая гипотеза для t-статистики и, соответственно, коэффициента регрессии проверяется с помощью таблиц распределения Стьюдента. В данном случае ясно без таблиц, по общему порядку цифр, что коэффициент при *GNP*, равный 0,017, статистически значим (так как $t_{GNP} = 4{,}25$), а коэффициент при *RSR*, равный (-0,411), статистически незначим. Его t-статистика $t_{RSR} = -0{,}411/0{,}947 \approx -0{,}434$ слишком мала по абсолютной величине. Если уточнить по таблицам, уровень значимости здесь составляет примерно . Следовательно, если в действительности (для генеральной совокупности) этот коэффициент равен нулю, то вполне вероятно (с вероятностью 2/3) для данного размера выборки (60 наблюдений) при двух объясняющих переменных получить такую (-0,434) или большую по модулю t-статистику данного коэффициента регрессии. Для оценки значимости коэффициента регрессии можно воспользоваться следующим грубым правилом: если абсолютная величина коэффициента меньше, чем его стандартная ошибка, то он статистически незначим (если нет мультиколлинеарности, или коррелированности объясняющих переменных, о которой речь пойдет позже). В данном случае это правило срабатывает, и на следующем шаге мы заменим переменную *RSR*.

Теперь рассчитаем *F*-статистику оцененного уравнения:

$$F = \frac{R^2}{1 - R^2} \frac{(n - m - 1)}{m} = \frac{0{,}29}{0{,}71} \frac{57}{2} \approx 11{,}6. \qquad (3)$$

По таблице распределения Фишера с (2; 57) степенями свободы находим, что критическое значение F равно 3,16 при 5 %-ном уровне значимости и 5,0 при 1%-ном. Таким образом, гипотеза о равенстве нулю одновременно всех коэффициентов регрессии заведомо отвергается (что, впрочем, ясно и из того, что коэффициент при *GNP* уже до этого получился значимым). Итак, даже небольшая величина $R^2 = 0{,}29$ при довольно большом числе наблюдений дала значимую величину *F*-статистики. В то же время если величина R^2 рассматривается как самостоятельный критерий качества регрессии (а не только как средство проверки нулевой гипотезы для всех коэффициентов одновременно), позволяющий оценить его в сравнении с качеством линии $y = \bar{y}$, то значение $R^2 = 0{,}29$ вряд ли можно считать хорошим. Это говорит о необходимости дальнейшего поиска объясняющих переменных для показателя *RNX*.

Для оценки качества множественной линейной регрессии и проверки наличия предполагавшихся свойств отклонений e_i нужна также статистика Дарбина-Уотсона *DW*. В рассматриваемом примере она равна 0,43. Невооруженным взглядом видна положительная автокорреляция e_i: *DW* близка к нулю. Проверим статистику *DW* по таблице для $n = 60$; $m = 2$ при уровне значимости 5%. Критические значения $d_l = 1{,}44$; $d_u = 1{,}57$. Поскольку $DW = 0{,}43 < 1{,}44 = d_l$, принимается гипотеза о наличии положительной автокорреляции остатков первого порядка. Таким образом, значение статистики Дарбина-Уотсона говорит о том, что оцениваемая зависимость имеет другой вид: действовали какие-то неучтенные факторы либо сама формула связи была нелинейной. Заметим, что если оцениваются регрессионные связи макроэкономических показателей по временным рядам наблюдений за столь длительный период времени, то статистика DW чаще всего оказывается близкой к нулю. Практически всегда какие-то факторы действуют на протяжении некоторых периодов времени, "уводя" зависимую переменную вверх или вниз от линии (или поверхности) регрессии. Идентификация таких

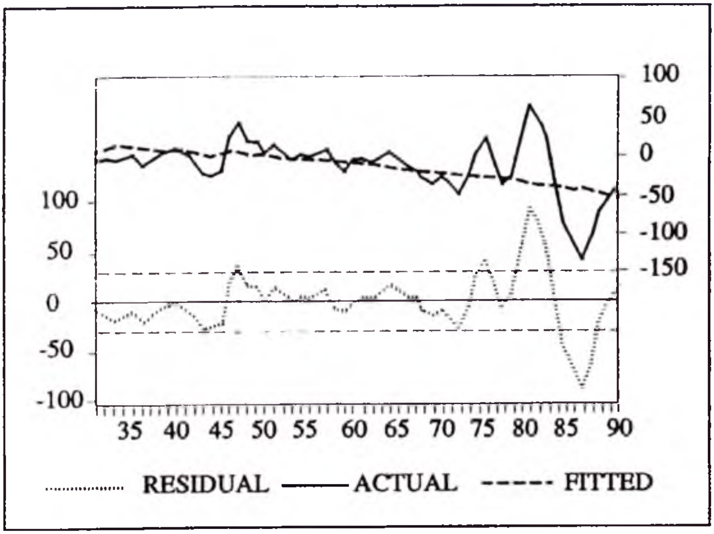


Рисунок 18.1. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины *RNX*; отклонения от линии регрессии (*Residual*) (США, 1931-1990 гг., уравнение: $RNX = 21{,}1 - 0{,}017\ GNP - 0{,}411\ RSR$)

факторов и определение периодов их действия является важной задачей эконометрики.

Указанные недостатки оцененного уравнения регрессии проявляются и на графике. На рис. 18.1, где показаны зависимости от времени действительных и рассчитанных по уравнению регрессии значений *RNX*, а также отклонений первых от вторых, можно видеть, что оцененное уравнение не описывает колебаний переменной *RNX*, а объясняет лишь ее общий тренд. Здесь же видно, что отклонения зависимой переменной от линии регрессии не являются независимыми и, кроме того, дисперсия их для разных периодов не постоянна.

Воздействие процентной ставки на величину чистого экспорта происходит с определенным временным запаздыванием (лагом). Заключаемые контракты ориентируются на текущий валютный курс (который, в свою очередь, с некоторой задержкой реагирует на изменения процентной ставки), а их исполнение обычно происходит лишь через несколько месяцев. Поэтому естественно в качестве первого шага в развитии модели чистого экспорта не исключать объясняющую переменную *RSR*, а ввести ее с лагом в один год, то есть заменить *RSR* на *RSR*(-1). В результате расчетный период сокращается на одну точку, то есть охватывает 1932-1990 гг. Получается следующее уравнение регрессии:

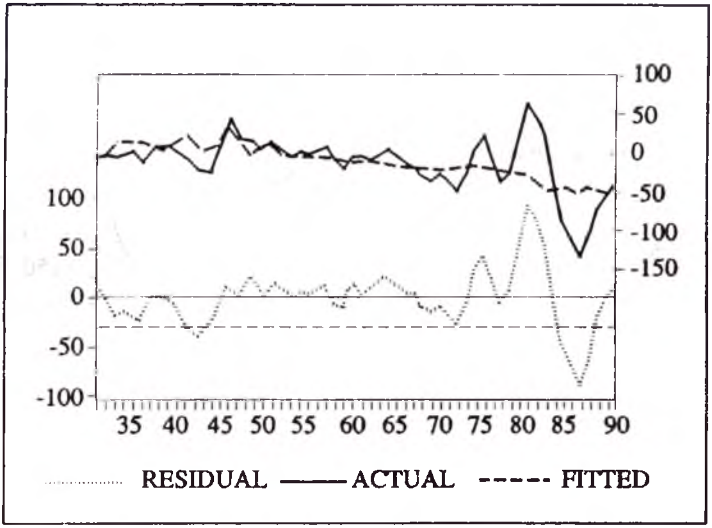


Рисунок 18.2. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины *RNX*; отклонения от линии регрессии (*Residual*) (США, 1932-1990 гг., уравнение: $RNX = 18{,}9 - 0{,}015\ GNP - 2{,}086\ RSR(-1)$)

$$RNX = \underset{(8{,}31)}{18{,}9} - \underset{(0{,}004)}{0{,}015}\,GNP - \underset{(0{,}911)}{2{,}086}\,RSR(-1) \qquad (4)$$

(в скобках приведены стандартные ошибки)

$R^2 = 0{,}35$; $DW = 0{,}46$.

Здесь обе объясняющие переменные статистически значимы; их t-статистики превышают по модулю 2. Однако обобщающие показатели качества модели R^2 и DW по сравнению с уравнением (2) существенно не улучшились. На графике (рис. 18.2) можно видеть, что в некоторые периоды, особенно во второй половине 1940-х - первой половине 1950-х годов, эта модель описывает уже не только общий тренд величины RNX, но и отклонения от этого тренда. В то же время она, безусловно, не подходит для всего периода 1931-1990 гг. Мы отметим это для дальнейшего, когда будем строить отдельные модели для различных частей рассматриваемого периода, но пока продолжим поиск единой модели для всего периода.

18.2. Уточнение состава объясняющих переменных в регрессионной модели

В качестве очередного шага в развитии модели мы заменим переменную *RSR* в уравнении регрессии (2) на переменную *ER (Real Exchange Rate)* - реальный курс доллара США по отношению к "корзине" основных иностранных валют. Причиной такой замены является то, что процентные ставки влияют на реальный экспорт и импорт косвенно, через валютный курс: чтобы принести больший процент в какой-то стране, деньги должны быть обменяны на валюту этой страны. Для обеспечения такого обмена увеличивается экспорт в эту страну и сокращается импорт. Кроме того, существует непосредственное влияние обменного курса на экспорт и импорт, не связанное прямо с процентной ставкой. Чем выше курс отечественной валюты, тем труднее экспортировать и легче импортировать, безотносительно к тому, где и под какой процент занимаются деньги для внешнеторговых операций и в банк какой страны они кладутся. Таким образом, для движения денег в международной торговле процентные ставки не так важны, как обменные курсы валют. Поэтому можно ожидать, что обменный курс окажется более значимой переменной в функции чистого экспорта, чем та или иная процентная ставка (кратко-, средне- или долгосрочная, номинальная или реальная). Итак, было оценено следующее уравнение:

$$\underset{(36,4)}{RNX = 94,2} - \underset{(0,004)}{0,018 \cdot GNP} - \underset{(0,309)}{0,631 \cdot ER} \qquad (5)$$

(в скобках приведены стандартные ошибки)
$R^2 = 0,33$; $DW = 0,41$.

Коэффициент при переменной *ER* в уравнении (5) примерно того же порядка, что и у *RSR* в (2), но его стандартная ошибка здесь в три раза меньше. Абсолютная величина t-статистики $t = 0,631/0,309$ 2,04 довольно значимо подтверждает (доверительная вероятность 95,4%), что коэффициент регрессии не равен нулю и что существует статистическая зависимость между обменным курсом и чистым экспортом. На графике (рис. 18.3) можно видеть, что рассчитанная по уравнению регрессии величина *RNX* отражает не только общий тренд действительного показателя, но также и (в некоторой степени) отклонения от этого тренда. Однако в целом поведение показателя *RNX* описано здесь ненамного лучше, чем в (2): доля объясненной дисперсии *RNX* осталась примерно такой же, как в (2) или в (4): ($R^2 = 0,33 = 33\%$). Статистика Дарбина-Уотсона *DW* также осталась практически на прежнем, слишком близком к нулю (то есть свидетельствующем о положительной автокорреляции остат-

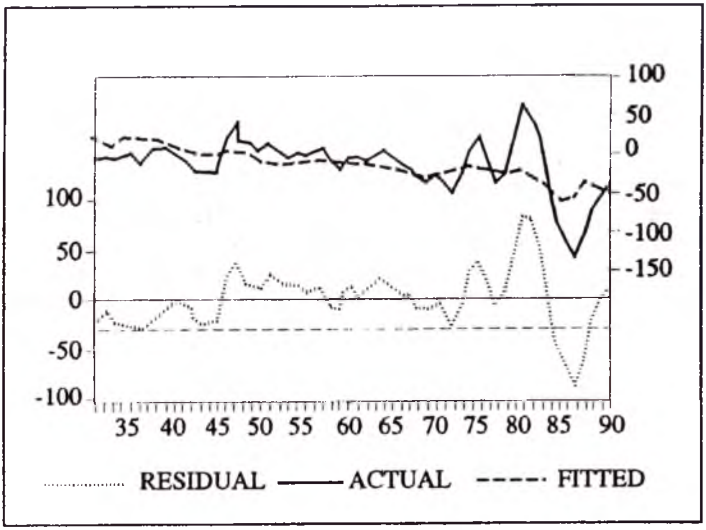


Рисунок 18.3. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины *RNX*; отклонения от линии регрессии (*Residual*) (США, 1931-1990 гг., уравнение: $RNX = 94{,}2 - 0{,}018\ GNP - 0{,}631\ ER$)

ков) уровне. Для улучшения качества зависимости целесообразно поискать дополнительный объясняющий фактор.

Введем в уравнение регрессии как дополнительную объясняющую переменную валютный курс предыдущего года $ER(-1)$. Причиной для этого служит то, что экспортно-импортные контракты в среднем подписываются примерно за полгода до их фактического осуществления. Значит, и в рассматриваемой нами функции зависимости чистого экспорта должен быть примерно полугодовой временной лаг между экспортно-импортными потоками и влияющими на них показателями, в частности валютным курсом. В этом случае, при дискретной годовой единице времени, в качестве объясняющих переменных должны присутствовать ER и $ER(-1)$.

Итак, следующим за период 1932-1990 гг. оценено следующее уравнение:

$$RNX = \underset{(34,4)}{143{,}0} - \underset{(0,003)}{0{,}018}\,GNP - \underset{(0,47)}{1{,}92}\,ER(-1) + \underset{(0,46)}{0{,}84}\,ER \qquad (6)$$

(в скобках приведены стандартные ошибки)

$R^2 = 0{,}50$; $DW = 0{,}54$.

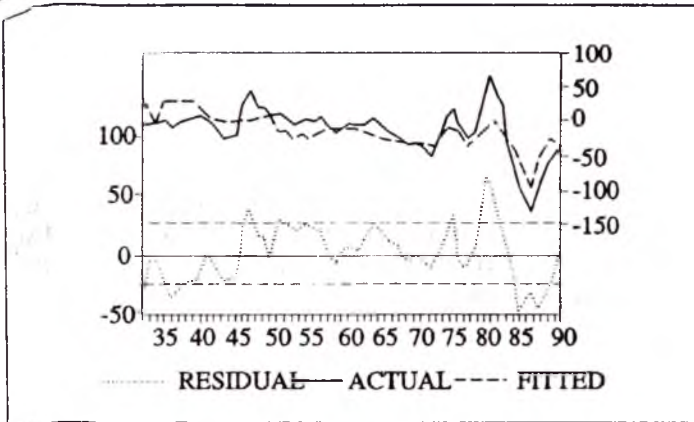


Рисунок 18.4. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины *RNX*; отклонения от линии регрессии (*Residual*) (США, 1932-1990, уравнение: $RNX = 143{,}0 - 0{,}018\,GNP - 1{,}92\,ER(-1) + 0{,}84\,\Delta ER$)

В уравнении (6) все коэффициенты статистически значимы ($t>2$), доля объясненной дисперсии переменной *RNX* выросла до 50%. Требует здесь пояснений положительный знак коэффициента при переменной *ER*, поскольку рост валютного курса должен приводить к сокращению, а не к увеличению чистого экспорта. Для этого перепишем уравнение регрессии в следующей форме:

$$RNX = \underset{(34,4)}{143{,}0} - \underset{(0,003)}{0{,}018}\,GNP - \underset{(0,29)}{1{,}08}\,ER + \underset{(0,47)}{1{,}92}\,\Delta ER \qquad (7)$$

(в скобках приведены стандартные ошибки)

где $\Delta ER = ER - ER(-1)$.

В уравнении (7) у переменной *ER* уже отрицательный знак, а положительный знак у переменной *ER* означает, что чем больше эта величина, тем (при данной *ER*) меньше *ER*(-1), с которой величина чистого экспорта связана отрицательной зависимостью. Можно оценить любую из этих зависимостей (включая также зависимость 8), получив эквивалентные результаты с теми же значениями стандартных ошибок коэффициентов, и трансформировать ее затем в любую другую зависимость.

$$RNX = \underset{(34,4)}{143{,}0} - \underset{(0,003)}{0{,}018}\,GNP - \underset{(0,29)}{1{,}08}\,ER(-1) + \underset{(0,46)}{0{,}84}\,\Delta ER \qquad (8)$$

(в скобках приведены стандартные ошибки).

Обратите также внимание на соотношения коэффициентов и стандартных ошибок в уравнениях (6)-(8). Значения коэффициентов меняются в зависимости от спецификации, а значения стандартных ошибок сохраняются. Таким образом, t-статистики различны, и можно подобрать такую спецификацию из нескольких эквивалентных друг другу, которая дает наибольшие по модулю величины t-статистик. С этой точки зрения уравнение (7) лучше, чем (6) или (8), так как у него все t-статистики превышают по модулю 3. Выбор уравнения (7) можно сделать и на основе требования наиболее наглядной содержательной интерпретации.

Добавим теперь вновь к нашей модели объясняющую переменную $RSR(-1)$, уже показавшую свою существенность для одного из промежутков рассматриваемого периода времени. Получаем уравнение

$$\underset{}{RNX} = \underset{(33{,}8)}{117{,}9} - \underset{(0{,}003)}{0{,}014}\,GNP - \underset{(0{,}29)}{0{,}90}\,ER + \underset{(0{,}45)}{2{,}09}\,\Delta ER - \underset{(0{,}81)}{2{,}18}\,RSR(-1) \qquad (9)$$

(в скобках приведены стандартные ошибки).

$R^2 = 0{,}56;\ DW = 0{,}68.$

В уравнении (9) все переменные статистически значимы, что связано с тем, что вновь введенная переменная $RSR(-1)$ объясняет поведение зависимой переменной RNX на временном диапазоне 1946-1955 гг., где не прослеживается влияния других переменных. В то же время, доля объясненной дисперсии зависимой переменной не возросла существенно по сравнению с (6), а статистика Дарбина-Уотсона остается близкой к нулю, что говорит о положительной автокорреляции остатков e_t и требует уточнения вида зависимости или состава объясняющих переменных. Те же выводы могут быть сделаны и на основании графика (рис. 5).

Из графика видно, что колебания рассчитанных по уравнению регрессии величин RNX частично соответствуют колебаниям фактических величин RNX в периоды примерно с середины 1940-х до начала 1950-х и с начала 1970-х годов. До середины 1940-х годов и в 1950-е - 1960-е годы модель позволяет описать лишь общий тренд показателя RNX, что в данном случае совершенно недостаточно. Разное поведение отклонений e_t на различных промежутках рассматриваемого периода и разный состав объясняющих факторов говорят о том, что спецификация модели для этих промежутков должна быть различной.

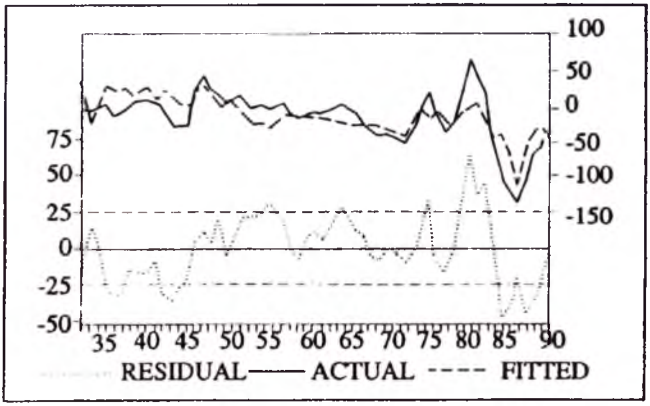


Рисунок 18.5. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины *RNX*; отклонения от линии регрессии (*Residual*) (США, 1932-1990 гг., уравнение: $RNX = 117{,}9 - 0{,}014 \cdot GNP - 0{,}90 \cdot ER + 2{,}09 \cdot ER - 2{,}18 \cdot RSR(-1)$)

18.3. Корректировка интервала оценивания линейной регрессионной модели

Из анализа оцененных уравнений регрессии для модели чистого экспорта можно сделать вывод, что изменения валютного курса стали определять величину реального чистого экспорта США лишь на рубеже 1960-х-1970-х годов. Для столь резкого изменения взаимодействия макроэкономических показателей должна быть какая-то объективная причина, выражающаяся в изменении экономической политики или внешнем "шоке". В данном случае объяснением может служить тот факт, что с 1971 года был введен плавающий курс доллара по отношению к валютам других развитых стран. До этого номинальный курс фиксировался, что искажало реальный курс и его воздействие на экспортно-импортные потоки. Лишь во второй половине 1940-х годов роль валютного курса в объяснении поведения показателя *RNX* играла процентная ставка *RSR*(-1).

В целом, говоря о разделении временного интервала на части, отметим, что оно необходимо в тех случаях, когда значения параметров $\{\alpha_i\}$ менялись во времени (что нарушало предпосылку модели линейной регрессии об их неизменности). Если изменялись они более или менее скачкообразно, то, разделяя временной интервал моментами таких "скачков", можно разбить его на несколько интервалов, на каждом из которых предпосылки модели выполнялись. Для проверки статистической значимости различия коэффициентов на отдельных отрезках рассматриваемого временного интервала можно использовать *t*-статистику, о чем говорилось в предыдущей главе.

Итак, на следующем шаге мы меняем стратегию: вместо введения новых объясняющих переменных разобьем рассматриваемый временной интервал 1931-1990 гг. на части и будем дальше совершенствовать каждое уравнение регрессии для того временного промежутка, на котором оно работает лучше.

Вначале выберем период 1946-1962 годов и оценим для него уравнение типа (4):

$$RNX = \underset{(14,1)}{49{,}9} - \underset{(0{,}010)}{0{,}032} \cdot GNP - \underset{(0{,}485)}{1{,}538} \cdot RSR(-1) \qquad (10)$$

(в скобках приведены стандартные ошибки)

$R^2 = 0{,}83$; $DW = 2{,}10$.

Уравнение (10) дает пример хорошей по всем параметрам эконометрической модели. Коэффициенты регрессии значимы; их *t*-статистики по модулю превышают 3. Доля объясненной дисперсии зависимой переменной (R^2) составляет 83%, а статистика Дарбина-Уотсона очень близка к 2, что позволяет с высокой степенью уверенности принять гипотезу об отсутствии автокорреляции остатков первого порядка. В то же время график (рис. 18.6) показывает, что данная модель действительно хороша лишь для первой половины расчетного периода; во второй его половине модель описывает лишь общий тренд показателя *RNX* и не отражает отклонений от этого тренда.

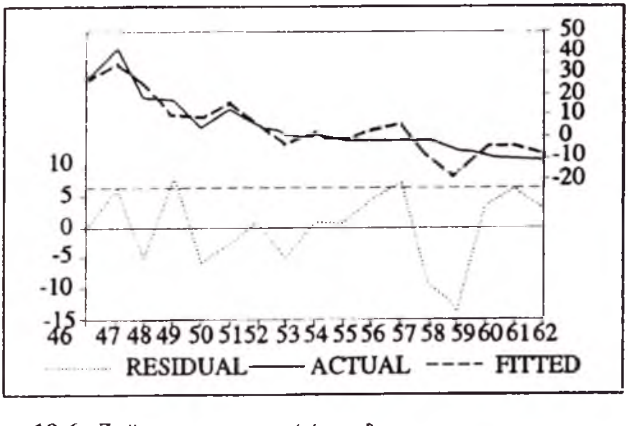


Рисунок 18.6. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины *RNX*; отклонения от линии регрессии (*Residual*) (США, 1946-1962 гг., уравнение: $RNX = 49{,}9 - 0{,}032 \cdot GNP - 1{,}538 \cdot RSR(-1)$)

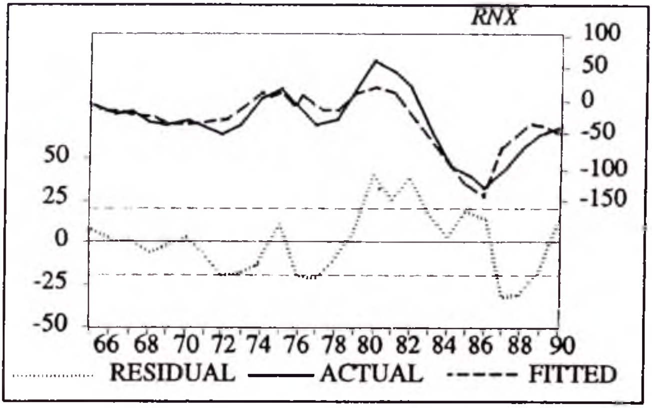


Рисунок 18.7. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины *RNX*; отклонения от линии регрессии (*Residual*) (США, 1965-1990 гг., уравнение: $RNX = 393{,}7 - 0{,}046 \cdot GNP - 2{,}58 \cdot ER(-1)$)

Теперь выберем период 1965-1990 гг. (чтобы иметь несколько дополнительных наблюдений перед 1971 годом) и начнем оценивание с уравнения типа (5), включив в него временной лаг в 1 год. Получаем

$$RNX = \underset{(41{,}0)}{393{,}7} - \underset{(0{,}007)}{0{,}046} \cdot GNP - \underset{(0{,}28)}{2{,}58} \cdot ER(-1) \qquad (11)$$

(в скобках приведены стандартные ошибки)

$R^2 = 0{,}82$; $DW = 0{,}89$.

Это уравнение намного лучше, чем (5). Все коэффициенты статистически значимы, их коэффициенты по абсолютной величине в 7-10 раз превышают свои стандартные ошибки. Уравнение соответствует макроэкономической теории, говорящей об отрицательной зависимости величины реального чистого экспорта от реального ВНП и валютного курса. Взглянув на рис. 18.7, можно отметить, что рассчитанные по уравнению регрессии величины ВНП за 1965-1990 гг. очень близки к фактическим. Единственной проблемой является то, что статистика Дарбина-Уотсона существенно меньше двух, - таким образом, можно попытаться улучшить это уравнение. При этом мы надеемся избавиться от автокорреляции остатков (то есть, получить более близкую к двум DW) и, возможно, увеличить долю объясненной дисперсии RNX, то есть R^2.

18.4. Мультиколлинеарность

Теперь мы предпримем для иллюстрации шаг, который впоследствии окажется ошибочным, но поможет при этом показать очень важное явление в оценивании множественной регрессии - мультиколлинеарность. Мультиколлинеарность - это коррелированность двух или нескольких объясняющих переменных в уравнении регрессии. Проблема мультиколлинеарности возникает только для случая множественной регрессии, поскольку в парной регрессии лишь одна объясняющая переменная. Оценка коэффициента регрессии может оказаться незначимой не только из-за несущественности данного фактора, но и из-за того, что трудно разграничить воздействие на зависимую переменную двух или нескольких факторов. Это бывает в том случае, когда какие-то факторы линейно связаны между собой (коррелированы) и меняются синхронно. Связь зависимой переменной с изменениями каждого из них можно определить, только если в число объясняющих переменных включается лишь один из этих факторов.

Природа мультиколлинеарности нагляднее всего может быть продемонстрирована на примере совершенной мультиколлинеарности, то есть строгой линейной связи между объясняющими переменными. Например, если в уравнении

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \qquad (12)$$

объясняющие переменные x_1 и x_2 связаны линейным соотношением $x_2 = \lambda x_1$, то исходное уравнение сводится к уравнению простой линейной регрессии

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \lambda x_1 + \epsilon = \beta_0 + \beta_1' x_1 + \epsilon, \qquad (13)$$

в котором могут быть получены оценки коэффициентов β_0 и $\beta_1' = \beta_1 + \lambda \beta_2$. Последнее уравнение представляет собой одно уравнение с двумя неизвестными β_1 и β_2, которые найдены по отдельности, естественно быть не могут. Таким образом, совершенная мультиколлинеарность не позволяет определить коэффициенты регрессии (в данном примере β_1 и β_2) и разделить вклады переменных x_1 и x_2 в объяснение поведения переменной y.

Несовершенная мультиколлинеарность, то есть стохастическая связь переменных x_1 и x_2, характеризуется величиной коэффициента корреляции с между ними. Чем ближе по абсолютной величине значение коэффициента корреляции к единице, тем ближе мультиколлинеарность к совершенной и тем труднее разделить влияния объясняющих переменных x_1 и x_2 на поведение переменной y и тем менее надежными будут оценки коэффициентов регрессии при этих переменных.

Предположим, что чистый экспорт зависит от потребления: поскольку существенная часть импортируемой продукции потребляется, для этого есть основания. Для этого введем в последнее уравнение дополнительную объясняющую переменную: величину реальных потребительских расходов *CONS (Real Consumption Expenditures)*. Получаем уравнение

$$\underset{(70{,}1)}{RNX = 406{,}6} - \underset{(0{,}110)}{0{,}07 \cdot GNP} - \underset{(0{,}31)}{2{,}61 \cdot ER(-1)} + \underset{(0{,}152)}{0{,}035 \cdot CONS} \qquad (14)$$

(в скобках приведены стандартные ошибки)

$R^2 = 0{,}82$; $DW = 0{,}87$.

Мы видим здесь, что R^2 и DW не стали лучше - переменная $CONS$ не дает дополнительной информации для объяснения поведения RNX. Однако здесь сразу два коэффициента регрессии оказались статистически незначимыми - коэффициент при переменной потребления $CONS$ (что понятно) и коэффициент при переменной GNP (что требует объяснения). Проблема здесь состоит в том, что переменные GNP и $CONS$ сильно коррелированы между собой - за период 1965-1990 гг. их коэффициент корреляции равен 0,9978. Итак, даже если существует сильное влияние на RNX обеих переменных GNP и $CONS$, мы не можем разделить это влияние при оценивании регрессии по данным нашей выборки. Поэтому мы исключаем из числа объясняющих только что введенную переменную $CONS$.

В общем случае, если при оценке уравнения регрессии несколько факторов оказались незначимыми, то нужно выяснить, нет ли среди них сильно коррелированных между собой. Для этого распечатывается корреляционная матрица (это предусмотрено стандартными статистическими программными пакетами), и проверяется статистическая значимость коэффициентов парной корреляции. При наличии корреляции один из пары связанных между собой факторов исключается, либо в качестве объясняющего фактора берется какая-то их функция. Если же незначимым оказался только один фактор, то можно его исключить или заменить другим (хотя, возможно, на каком-то более коротком промежутке времени данный фактор оказался бы значимым).

18.5. Корректировка модели чистого экспорта

Следующий шаг в развитии модели функции чистого экспорта достаточно очевиден - мы добавляем объясняющую переменную ΔER, которая уже показала свою значимость в функции чистого экспорта за весь период 1931 - 1990 гг. Получаем уравнение

$$RNX = \underset{(42,3)}{370,5} - \underset{(0,007)}{0,044}\cdot GNP - \underset{(0,29)}{2,42}\cdot ER(-1) + \underset{(0,42)}{0,67}\cdot \Delta ER \quad (15)$$

(в скобках приведены стандартные ошибки)

$R^2 = 0,84$; $DW = 1,32$.

Здесь коэффициент при ΔER не слишком значим, но если переписать то же уравнение по-другому, он оказывается вполне статистически значимым:

$$RNX = \underset{(42,3)}{370,5} - \underset{(0,007)}{0,044}\cdot GNP - \underset{(0,29)}{2,42}\cdot ER + \underset{(0,42)}{3,10}\cdot \Delta ER \quad (16)$$

(в скобках приведены стандартные ошибки).

Статистика Дарбина-Уотсона здесь намного лучше, чем у предыдущих уравнений, но еще остаются возможности улучшения как ее, так и доли объясненной дисперсии *RNX*. Введение объясняющей переменной прироста ВНП $\Delta GNP = GNP - GNP(-1)$ помогает увеличить значение R^2:

$$RNX = \underset{(34,9)}{339,0} - \underset{(0,005)}{0,038}\cdot GNP - \underset{(0,048)}{0,177}\cdot \Delta GNP - \underset{(0,24)}{2,15}\cdot ER + \underset{(0,35)}{2,80}\cdot \Delta ER \quad (17)$$

(в скобках приведены стандартные ошибки)

$R^2=0,90$; $DW=1,12$.

Здесь коэффициент при переменной ΔGNP отрицателен, как и коэффициент при *GNP*. Для последнего причина ясна: рост дохода в первую очередь вызывает рост импорта. Отрицательный знак коэффициента при ΔGNP может быть объяснен следующим образом. При данном значении текущего *GNP* большая величина *GNP* означает больший его прирост, что учитывается потребителем при прогнозировании прироста дохода в будущем году. Если он ожидает, что в следующем году его доход существенно возрастет, то он может увеличить потребление импортных товаров уже в текущем году. В результате сокращается чистый экспорт. Здесь можно видеть разницу в воздействии на чистый экспорт приростов валютного курса и объема ВНП: важными оказались валютный курс предыдущего года и ожидаемый доход следующего года. Поэтому, хотя знаки коэффициентов переменных *ER* и *GNP* оба отрицательны, знаки при величинах прироста этих показателей в уравнении регрессии различны.

Полученное уравнение приемлемо по всем параметрам, лишь статистика Дарбина-Уотсона говорит о возможном наличии некоторой автокорреляции остатков. Возможность добавления новых объясняющих переменных здесь уже практически исчерпанна, и можно считать, что остающаяся автокорреляция остатков обусловлена внутренними свойствами рядов e_i. В этом случае, например, можно использовать авторегрессионное преобразование первого порядка $AR(1)$, смысл которого в учете линейной регрессионной связи соседних отклонений e_i. Это будет сделано в главе 19.

18.6. Простейшие методы линеаризации

В процессе построения и развития модели функции чистого экспорта мы рассмотрели ряд основных подходов к улучшению статистического качества модели. До сих пор мы говорили лишь о линейной регрессионной модели, однако нередко связь между экономическими переменными существенно нелинейна. В заключение этой главы рассмотрим некоторые методы сведения нелинейной модели к линейной, или ее линеаризации. Сделаем это на примере построения макроэкономических производственных функций.

Предположим, что для некоторой модели линейная спецификация не дала приемлемых результатов, и из анализа различных статистик и графиков мы установили, что связь переменных нелинейна. Это означает, что нужно оценить уравнение нелинейной регрессии. Для оценки нелинейной регрессии существуют различные пути. Во-первых, существуют методы и алгоритмы оценивания нелинейных зависимостей: предложенная из априорных соображений формула оценивается, например, методом наименьших квадратов. Здесь, так как речь идет о линейной регрессии, мы эти методы рассматривать не будем.

Если нелинейная зависимость может быть записана в виде суммы функций от неизвестных x_i (например, $y = a + bx_1 + cx_1^2 + hx_2$), то можно построить новые ряды данных (для примера в скобках - ряд данных x_1^2) и оценить с ними линейную регрессию. Наиболее распространенные виды функций и преобразований данных, необходимые для построения нужного набора новых переменных, обычно заложены в прикладные регрессионные пакеты. Пусть, например, требуется оценить параметры производственной функции Кобба-Дугласа $Y = AK^{\alpha}L^{\beta}$. Для линеаризации прологарифмируем обе части:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L. \tag{18}$$

Полученная формула линейна относительно логарифмов выпуска Y, капитала K и труда L, и она может быть оценена как множественная линейная регрессия. Более сложные формулы (например, функцию CES

$$Y = A\,(ux_1^{-b}+(1-u)x_2^{-b})^{-n/b}) \tag{19}$$

можно оценить путем разложения в ряд. Как известно, любая дифференцируемая функция может быть разложена в ряд по степеням независимой переменной x в окрестности любой точки. Затем оставляются несколько наиболее важных членов ряда (остальные отбрасываются), и по ним оценивается линейная регрессия.

Если нужно оценить ПФ Кобба-Дугласа с $\alpha+\beta=1$, то делается следующее преобразование:

$$Y = AK^{\alpha}L^{\alpha-1} \Rightarrow \frac{Y}{L} = A\left[\frac{K}{L}\right]^{\alpha} \Rightarrow \ln\left[\frac{Y}{L}\right] = \ln A + \ln\left[\frac{K}{L}\right]. \qquad (20)$$

Далее оценивается парная линейная регрессия логарифма производительности труда $\left[\frac{Y}{L}\right]$ от логарифма капиталовооруженности $\left[\frac{K}{L}\right]$.

Нужно иметь в виду, что если с формулой связи делаются какие-то преобразования, то меняются свойства ошибок ϵ_i. Если для них предполагалось нормальное распределение с нулевым математическим ожиданием, то после, например, логарифмирования правой части оно уже таким не будет. Это серьезная проблема, изучаемая эконометрикой. Мы на ней останавливаться не будем, просто отметив ее наличие. Для простоты будем считать, что (там, где это возможно) отклонения ϵ_i обладают нужными свойствами именно у итоговой, линеаризованной зависимости.

Если зависимость оценивается по данным временных рядов, то часть тренда зависимой переменной может объясниться действовавшими во времени факторами, которые в совокупности могут учитываться просто включением в уравнение некоторой зависимости от времени. Такая зависимость может быть, например, линейной или экспоненциальной (изменение с постоянным темпом). В частности, ПФ Кобба-Дугласа может учитывать нейтральный технический прогресс с помощью множителя $e^{\gamma t}$:

$$Y = AK^{\alpha}L^{\beta}e^{\gamma t}. \qquad (21)$$

После линеаризации эта формула становится следующей: $\ln Y = \ln A + \alpha\ln K + \beta\ln L + \gamma t$ и может быть оценена с помощью модели линейной регрессии.

Вопросы к главе 18

1. В каких случаях необходимо уточнение линейной регрессионной модели и как оно осуществляется? Для пунктов 2-5 приведите соответствующие примеры с моделью чистого экспорта.
2. Когда необходимо выведение из рассмотрения незначимых объясняющих переменных и добавление новых переменных?
3. В каких случаях осуществляется разбиение временного интервала на части и оценка исходной или новой формулы регрессии на каждой из них?
4. Когда необходимо преобразование исходных данных с целью устранить их нежелательные свойства?
5. В каких случаях осуществляется построение нелинейных спецификаций уравнения регрессии с последующей их линеаризацией (или оценкой нелинейной регрессии)?
6. Какие объясняющие переменные включаются в модель чистого экспорта? Какие из них оказались значимыми и почему?
7. Объясните явление мультиколлинеарности. Что такое совершенная мультиколлинеарность?
8. Как устранить мультиколлинеарность? Поясните на примере модели чистого экспорта.
9. Как линеаризовать производственную функцию Кобба-Дугласа? Для чего это необходимо?
10. Какие проблемы спецификации ошибок возникают при линеаризации уравнения регрессии?

ГЛАВА 19
НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ ПРИКЛАДНЫЕ МЕТОДЫ ЭКОНОМЕТРИКИ

В предыдущих главах были рассмотрены модели парной и множественной линейной регрессии, а также задачи экономического анализа, решаемые с помощью этих моделей. Однако далеко не все задачи исследования взаимосвязей экономических переменных описываются обычной линейной регрессионной моделью. Во-первых, исходные данные могут не соответствовать тем или иным предпосылкам линейной регрессионной модели и требовать либо дополнительной обработки, либо иного модельного инструментария. Во-вторых, исследуемый процесс во многих случаях описывается не одним уравнением, а системой, где одни и те же переменные могут быть в одних случаях объясняющими, а в других - зависимыми. В-третьих, исследуемые взаимосвязи могут быть (и обычно являются) нелинейными, а процедура линеаризации не всегда легко осуществима и может приводить к искажениям. В-четвертых, структура описываемого процесса может обусловливать наличие различного рода связей между оцениваемыми коэффициентами регрессии, что также предполагает необходимость использования специальных методов. Настоящая глава посвящена обзору ситуаций, требующих выхода за рамки стандартной модели линейной регрессии, и подходов к их исследованию.

Наиболее распространенным в практике статистического оценивания параметров уравнений регрессии является метод наименьших квадратов. Этот метод основан на ряде предпосылок относительно природы данных и результатов построения модели. Основные из них - это четкое разделение исходных переменных на зависимые и независимые, некоррелированность факторов, входящих в уравнения, линейность связи, отсутствие автокорреляции остатков, равенство их математических ожиданий нулю и постоянная дисперсия. Эмпирические данные не всегда обладают такими характеристиками, т.е. предпосылки МНК нарушаются. Применение этого метода в чистом виде может привести к таким нежелательным результатам, как смещение оцениваемых параметров, снижение их состоятельности, устойчивости, а в некоторых случаях может и вовсе не дать решения. Для смягчения нежелательных эффектов при построении регрессионных уравнений, повышения адекватности моделей существует ряд усовершенствований МНК, которые применяются для данных нестандартной природы.

19.1. Взвешенный метод наименьших квадратов *(Weighted Least Squares, WLS)*

Одной из основных гипотез МНК является предположение о равенстве дисперсий отклонений e_i, т.е. их разброс вокруг среднего (нулевого) значения ряда должен быть величиной стабильной. Это свойство называется *гомоскедастичностью.* На практике дисперсии отклонений достаточно часто неодинаковы, то есть наблюдается *гетероскедастичность.* Это может быть следствием разных причин. Например, возможны ошибки в исходных данных. Случайные неточности в исходной информации, такие как ошибки в порядке чисел, могут оказать ощутимое влияние на результаты. Часто больший разброс отклонений ϵ_i наблюдается при больших значениях зависимой переменной (переменных). Если в данных содержится значительная ошибка, то, естественно, большим будет и отклонение модельного значения, рассчитанного по ошибочным данным. Для того, чтобы избавиться от этой ошибки нам нужно уменьшить вклад этих данных в результаты расчетов, задать для них меньший вес, чем для всех остальных. Эта идея реализована во взвешенном МНК.

Пусть на первом этапе оценена линейная регрессионная модель с помощью обычного МНК. Предположим, что остатки e_i независимы между собой, но имеют разные дисперсии (поскольку теоретические отклонения ϵ_i нельзя рассчитать, их обычно заменяют на фактические отклонения зависимой переменной от линии регрессии e_i, для которых формулируются те же исходные требования, что и для ϵ_i). В этом случае квадратную матрицу ковариаций $\mathrm{cov}(e_i,e_j)$ можно представить в виде:

$$V = \begin{vmatrix} s_1^2 & 0 & 0 & \dots & 0 \\ 0 & s_2^2 & 0 & \dots & 0 \\ 0 & 0 & s_3^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & s_n^2 \end{vmatrix},$$

где $\mathrm{cov}(e_i,e_j)=0$ при $i \neq j$; $\mathrm{cov}(e_i,e_i)=s_i^2$; n - длина рассматриваемого временного ряда.

Если величины s_i^2 известны, то далее можно применить взвешенный МНК, используя в качестве весов величины $\frac{1}{s_i^2}$ и минимизируя сумму

$$Q_1 = \sum_i \left\{ \frac{1}{s_i^2} \cdot (y_i - a - bx_i)^2 \right\}.$$

Формула Q_1 записана для парной регрессии; аналогичный вид она имеет и для множественной линейной регрессии. При использовании *WLS* оценки параметров не только получаются несмещенными (они будут таковыми и для обычного МНК), но и более точными (имеют меньшую дисперсию), чем невзвешенные оценки.

Проблема заключается в том, чтобы оценить величины s_i^2, поскольку заранее они обычно неизвестны. Поэтому, используя на первом этапе обычный МНК, нужно попробовать выяснить причину и характер различий дисперсий e_i. Для экономических данных, например, величина средней ошибки может быть пропорциональна абсолютному значению независимой переменной. Это можно проверить статистически и включить в расчет МНК веса, равные $\frac{1}{x_i^2}$. Существуют специальные критерии и процедуры проверки равенства дисперсий отклонений. Например, можно рассмотреть частное от деления сумм самых больших и самых маленьких квадратов отклонений, которое должно иметь распределение Фишера в случае гомоскедастичности.

Использование взвешенного метода в статистических пакетах, где предоставлена возможность задавать веса вручную, позволяет регулировать вклад тех или иных данных в результаты построения моделей. Это необходимо в тех случаях, когда мы априорно знаем о нетипичности какой-то части информации, т.е. на зависимую переменную оказывали влияние факторы, заведомо не включаемые в модель. В качестве примера такой ситуации можно привести случаи стихийных бедствий, засух. При анализе макроэкономических показателей (ВНП и др.) данные за эти годы будут не совсем типичными. В такой ситуации нужно попытаться исключить влияние этой части информации заданием весов. В разных статистических пакетах приводится возможный набор весов. Обычно это числа от 0 до 100. По умолчанию все данные учитываются с единичными весами. При указании веса меньше 1 мы снижаем вклад этих данных, а если задать вес больше единицы, то вклад этой части информации увеличится. Путем задания весового вектора мы можем не только уменьшить влияние каких - либо лет из набора данных, но и вовсе исключить его из анализа. Итак, ключевым моментом при применении этого метода является выбор весов. В первом приближении веса могут устанавливаться пропорционально ошибкам невзвешенной регрессии.

19.2. Системы одновременных уравнений

При статистическом моделировании экономических ситуаций часто необходимо построение систем уравнений, когда одни и те же переменные в различных регрессионных уравнениях могут одновременно выступать, с одной стороны, в роли результирующих, объясняемых переменных, а с другой стороны - в роли объясняющих переменных. Такие системы уравнений принято называть системами одновременных уравнений. При этом в соотношения могут входить переменные, относящиеся не только к текущему периоду t, но и к предшествующим периодам. Такие переменные называются лаговыми. Переменные за предшествующие годы обычно выступают в качестве объясняющих переменных.

В качестве иллюстрации приведем пример из экономики. Рассмотрим модель спроса и предложения. Как известно, спрос D на некоторый продукт зависит от его цены p. От этого же параметра, но с противоположным по знаку коэффициентом, зависит и предложение этого продукта. Силы рыночного механизма формируют цену таким образом, что спрос и предложение уравниваются. Нам нужно построить модель описанной ситуации. Для этого имеются данные об уровне равновесных цен и спросе (который равен предложению). Представленную ситуацию можно формализовать в виде следующей линейной модели:

1) $D_t = a_1 p_t + b_1 + e_t$ - (1)

спрос пропорционален цене с коэффициентом пропорциональности $a_1 < 0$, т.е. связь отрицательная;

2) $S_t = a_2 p_t + b_2 + e'_t$ - (2)

предложение пропорционально цене с коэффициентом пропорциональности $a_2 > 0$, т.е. связь положительная;

3) $S_t = D_t$. (3)

Здесь e_t, e'_t, (t=1,...,n) - ошибки модели, имеющие нулевое математическое ожидание.

Первые два из представленных уравнений, если их рассматривать отдельно, могут показаться вполне обычными. Мы можем определить коэффициенты регрессии для каждого из этих уравнений. Но в этом случае остается открытым вопрос о равенстве спроса и предложения, т.е. может не выполняться третье равенство, в котором спрос выступает в качестве зависимой переменной. Поэтому расчет параметров отдельных уравнений в такой ситуации теряет смысл.

Экономическая модель как система одновременных уравнений может быть представлена в структурной или в приведенной форме. В структурной форме ее уравнения имеют исходный вид, отражая непосредственные связи между переменными. Приведенная форма получается после решения модели относительно эндогенных (внутренних) переменных, то есть выражения этих переменных только через экзогенные (задаваемые извне) переменные и параметры модели. Например, в модели спроса и предложения эндогенными являются переменные p_t, S_t, D_t, ее параметры - a_1, a_2, b_1, b_2, а экзогенных переменных в ней нет. Таким образом, в приведенной форме переменные p_t, S_t, D_t должны выражаться только через параметры модели. Подставив D_t и S_t из (1) и (2) в (3), получаем

$$a_1 p_t + b_1 + e_t = a_2 p_t + b_2 + e_t \Rightarrow p_t = \frac{b_2 - b_1}{a_1 - a_2} + v_{1t};$$

$$D_t = a_1 \cdot \frac{b_2 - b_1}{a_1 - a_2} + b_1 + v_{2t};\ S_t = a_2 \cdot \frac{b_2 - b_1}{a_1 - a_2} + b_2 + v_{3t}.$$

Здесь v_{1t}, v_{2t}, v_{3t} - преобразованные отклонения. Мы можем оценить $\frac{b_2 - b_1}{a_1 - a_2}$ как среднее значение p_t (т.е. $\bar{p}$), а также $a_1\bar{p} + b_1 = D$, $a_2\bar{p} + b_2 = S$, но из этих трех соотношений невозможно рассчитать параметры первоначальной модели a_1, a_2, b_1 и b_2 (поскольку их четыре). Тем самым мы подошли к проблеме идентификации - оценке параметров структурной формы модели (в чем, собственно, и состоит наша задача) по параметрам приведенной формы. Параметры приведенной формы могут быть оценены обычным МНК, но по ним далеко не всегда может быть идентифицирована исходная модель (как, например, в описанном случае модели спроса и предложения). Для того чтобы структурная форма модели могла быть идентифицирована, вводят дополнительные предпосылки (например, о равенстве некоторых коэффициентов нулю или об их взаимосвязи между собой). Часто уже на этапе построения модели стараются выбрать такую ее форму, которая была бы идентифицируема. Такой, например, является треугольная форма модели:

$$y_1 = f(x);\ y_2 = f(y_1, x);\ \ldots;\ y_k = f(y_1, y_2, \ldots, y_{k-1}, x), \tag{4}$$

где x - вектор объясняющих переменных, y_i - i-я зависимая переменная. Нежелательна и сверхидентифицируемость модели, когда для параметров структурной формы получается слишком много соотношений из приведенной формы модели. В этом случае модель также нуждается в уточнении.

Для оценивания систем одновременных уравнений имеется ряд методов. В целом их можно разбить на две группы. К первой группе относятся методы, применяемые к каждому уравнению в отдельности. Вторая группа содержит методы, предназначенные для оценивания всей системы в целом. В пакете *TSP*, в частности, представлено по одному методу из каждой группы. Для оценивания отдельных уравнений можно применять *двухшаговый метод наименьших квадратов (Two-Stage Least Squares)*. Из второй группы методов в этом пакете реализован *трехшаговый метод наименьших квадратов (Three-Stage Least Squares)*.

Остановимся вначале на двухшаговом методе. Он применяется при наличии в оцениваемой модели лаговых переменных. Содержательный смысл двухшагового метода состоит в следующем. Как известно, МНК-оценки параметров уравнения равны $b = (X'X)^{-1} X'Y$, но лаговые значения y, используемые как объясняющие переменные (в этой формуле они являются частью матрицы X), заранее неизвестны. Поэтому для того, чтобы воспользоваться этой формулой, сначала, на первом шаге, определяются недостающие значения объясняемых переменных. Это в данном случае делается путем расчета МНК-оценок, т.е. строится регрессия, в которой в роли объясняемых переменных выступают только имеющиеся в исходной информации. После этого, когда исходные эмпирические данные дополнены рассчитанными значениями и сформирован полный набор данных, можно приступать к оценке искомых параметров.

Двухшаговый МНК применяется и при сверхидентифицируемости модели. В этом случае на первом шаге оцениваются параметры приведенной формы модели. С помощью уравнений приведенной формы, при заданных значениях объясняющих переменных, рассчитываются оценки зависимых переменных. Далее эти оценки подставляются в правые части уравнений модели в структурной форме, и вновь используется обычный МНК для оценки ее параметров.

Для оценки параметров всей системы уравнений в целом используется трехшаговый МНК. К его применению прибегают в тех случаях, когда переменные, объясняемые в одном уравнении, в другом выступают в роли объясняющих. Так было в нашем примере с моделью спроса и предложения, где спрос и предложение, с одной стороны, определяются рыночной ценой, а с другой стороны, предложение должно быть равно спросу. При расчете параметров таких моделей необходимо учитывать всю систему соотношений. В трехшаговом методе это реализуется в три этапа. Первые два из них похожи на двухшаговый метод, т.е. производится оценка параметров в уравнениях с лаговыми переменными. В нашем примере лаговые переменные в уравнения не включены, и на этом этапе будут рассчитываться обычные коэффициенты регрессии. После этого нам нужно увязать все уравнения системы между собой. В качестве меры связи здесь выступает матрица ковариаций ошибок моделей, т.е. чтобы оценить, насколько несвязанными получатся уравнения спроса и предложения при расчете их отдельно, нужно рассчитать ковариацию ошибок e и e'. Для увеличения этой связи на следующем этапе, при очередном расчете коэффициентов регрессии учитывается матрица ковариаций ошибок. Таким приемом достигается взаимосязанность всей системы уравнений.

19.3. Нелинейная регрессия

На практике часто встречается ситуация, когда априорно известен нелинейный характер зависимости между объясняемыми и объясняющими переменными. В этом случае функция f в уравнении $y=f(a,x)$ нелинейна (a - вектор параметров функции, которые нам нужно оценить). Например, вид зависимости между ценой и количеством товара в той же модели спроса и предложения: она не всегда предполагается линейной, как в нашем примере. Нелинейную функцию можно преобразовать в линейную, как это было сделано, например, логарифмированием с функцией Кобба-Дугласа. Однако не все функции поддаются такой непосредственной линеаризации. Любую дифференцируемую нужное число раз функцию можно разложить в функциональный ряд и затем оценить регрессию объясняемой переменной с членами этого ряда. Тем не менее такое разложение всегда осуществляется в окрестности определенной точки, и лишь в этой окрестности достаточно точно аппроксимирует оцениваемую функцию. В то же время оценить зависимость требуется обычно на более или менее значительном интервале, а не только в окрестности некоторой точки. При линеаризации функции или разложении её в ряд с целью оценки регрессии возникают и другие проблемы: искажение отклонений e и нарушение их первоначальных свойств, статистическая зависимость членов ряда между собой. Например, если оценивается формула

$$y = ax + bx^2 + e',$$

полученная путем линеаризации или разложения в ряд, то независимые переменные x и x^2 связаны между собой даже не статистически, но функционально. Если исходная ошибка e здесь связана с переменной x, то добавление x^2 приводит к появлению (с соответствующими коэффициентами) квадрата этой переменной и её удвоенного произведения с x, что искажает исходные предпосылки модели. Поэтому во многих случаях актуальна непосредственная оценка нелинейной формулы регрессии. Для этого можно воспользоваться нелинейным МНК. Идея МНК основана на том, чтобы минимизировать сумму квадратов отклонений расчетных значений от эмпирических, т.е. нужно оценить параметры a функции $f(a,x)$ таким образом, чтобы ошибки $e_i = y_i - f(a,x_i)$, точнее - их квадраты, по совокупности были минимальными. Для этого нужно решить задачу минимизации

$$F = \sum_i (y_i - f(a,x_i))^2 \rightarrow \min. \qquad (5)$$

Для решения этой задачи существует два пути. Во-первых, может быть осуществлена непосредственная минимизация функции F с помощью методов нелинейной оптимизации, позволяющих находить экстремумы выпуклых линий. Это, например, метод наискорейшего спуска, при использовании которого в некоторой исходной точке определяется антиградиент (направление наиболее быстрого убывания) функции F. Далее находится минимум F при движении в данном направлении, и в точке этого минимума снова определяется градиент. Процедура повторяется до тех пор, пока разница значений F на двух последовательных шагах не окажется меньше заданной малой величины. Другой путь состоит в решении системы нелинейных уравнений, которая получается из необходимых условий экстремума функции F. Эти условия - равенство нулю частных производных функции F по каждому из параметров a_j, т.е.

$$F'_{aj} = 0,$$

$j = 1,...,m$. Получается система уравнений

$$-2\sum_i (y_i - f(a,x_i)) \cdot f'_{aj}(a,x_i) = 0,\ j = 1,...,\ m, \qquad (6)$$

нелинейность которой обусловлена нелинейностью функции f относительно параметров a_j. Эта система уравнений может быть решена итерационными методами (когда последовательно находятся векторы параметров, все в меньшей степени нарушающие уравнения системы). Однако в общем случае решение такой системы не является более простым способом нахождения вектора a, чем непосредственная оптимизация методом наискорейшего спуска.

Существуют методы оценивания нелинейной регрессии, сочетающие непосредственную оптимизацию, использующую нахождение градиента, с разложением в функциональный ряд (ряд Тейлора) для последующей оценки линейной регрессии. Наиболее известен из них метод Марквардта, сочетающий в себе достоинства каждого из двух используемых методов.

При построении нелинейных уравнений более остро, чем в линейном случае, стоит проблема правильной оценки формы зависимости между переменными. Неточности при выборе формы оцениваемой функции существенно сказываются на качестве отдельных параметров уравнений регрессии и, соответственно, на адекватности всей модели в целом.

19.4. Авторегрессионное преобразование

Важной проблемой при оценивании регрессии является автокорреляция остатков e_i, которая говорит об отсутствии первоначально предполагавшейся их взаимной независимости. Автокорреляция остатков первого порядка, выявляемая с помощью статистики Дарбина-Уотсона, говорит о неверной спецификации уравнения либо о наличии неучтенных факторов. Естественно, для её устранения нужно попытаться выбрать более адекватную формулу зависимости, отыскать и включить важные неучтенные факторы или уточнить период оценивания регрессии. В некоторых случаях, однако, это не даст результата, а отклонения e_i просто связаны авторегрессионной зависимостью. Если это авторегрессия первого порядка, то её формула имеет вид $e_i=\rho e_{i-1} + u_i$ (ρ - коэффициент авторегрессии, $|\rho|<1$), и мы предполагаем, что остатки u_i в этой формуле обладают нужными свойствами, в частности - взаимно независимы. Оценив ρ, введем новые переменные $y'_i=y_i - \rho y_{i-1}$; $x'_i=x_i - \rho x_{i-1}$ (это преобразование называется авторегрессионным (AR), или преобразованием Бокса-Дженкинса). Пусть мы оцениваем первоначально формулу линейной регрессии $y_i= a + bx_i + e_i$. Тогда

$$y'_i = y_i - \rho y_{i-1} = a + bx_i + e_i - \rho(a + bx_{i-1} + e_{i-1}) =$$
$$= a(1 - \rho) + b(x_i - \rho x_{i-1}) + e_i - \rho e_{i-1} = a(1 - \rho)+bx'_i+u_i.$$

Если величины u_i действительно обладают нужными свойствами, то в линейной регрессионной зависимости $y'_i = a_1 + bx'_i + u_i$ автокорреляции остатков u_i уже не будет, и статистика DW окажется близкой к двум. Коэффициент b этой формулы принимается для исходной формулы $y = a+bx+e$ непосредственно, а коэффициент a рассчитывается по формуле $a = \dfrac{a_1}{1 - \rho}$.

Оценки коэффициентов a и b нужно сравнить с первоначальными оценками, полученными для расчета отклонений e_i. Если эти оценки совпадают, то процесс заканчивается; если нет - то при новых значениях a и b вновь рассчитываются отклонения e_i до тех пор, пока оценки a и b на двух соседних итерациях не совпадут с требуемой точностью.

В случае, когда остатки u_i также автокоррелированы, авторегрессионное преобразование может быть применено ещё раз. Это означает использование авторегрессионного преобразования более высокого порядка, которое заключается в оценке коэффициентов авторегрессии соответствующего порядка для отклонений e_i и использовании их для построения новых переменных. Такое преобразование вместо $AR(1)$ называется $AR(s)$ - если используется авторегрессия порядка s.

О целесообразности применения авторегрессионного преобразования говорит некоррелированность полученных отклонений u_i. Однако даже в этом случае истинной причиной первоначальной автокорреляции остатков может быть нелинейность формулы или неучтенный фактор. Мы же, вместо поиска этой причины, ликвидируем её бросающееся в глаза следствие. В этом - основной недостаток метода AR и содержательное ограничение для его применения.

Кроме авторегрессионного преобразования, для устранения автокорреляции остатков и уточнения формулы регрессионной зависимости может использоваться метод скользящих средних (*Moving Averages*, или MA). В этом случае считается, что отклонения от линии регрессии e_i описываются как скользящие средние случайных нормально распределенных ошибок ϵ_i : предполагается, что

$$e_i = \epsilon_i + \theta_1\epsilon_{i-1} + ... + \theta_q\epsilon_{i-q}. \tag{7}$$

Это формула для преобразования MA q-го порядка, или $MA(q)$; $MA(1)$, например, имеет вид $e_i = \epsilon_i + \theta_1\epsilon_{i-1}$. Параметры θ_i, как и в случае авторегрессионного преобразования, могут оцениваться итерационными методами.

Во многих случаях сочетание методов AR и MA позволяет решить проблему автокорреляции остатков даже при небольших s и q. Еще раз повторим, что адекватным такое решение проблемы является лишь в том случае, если автокорреляция остатков имеет собственные внутренние причины, а не вызвана наличием неучтенных (одного или нескольких) факторов.

Методы AR и MA могут использоваться в сочетании с переходом от объемных величин в модели к приростным, для которых статистическая взаимосвязь может быть более точной и явной. Модель, сочетающая все эти подходы, называется моделью *ARIMA (Autoregressive Integrated Moving Averages)*. В общем виде ее формулу можно записать так:

$$y^*_t = \rho_1 y^*_{t-1} + \rho_2 y^*_{t-2} + \rho_p y^*_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + ... + \theta_q\epsilon_{t-q}, \tag{8}$$

где $\{\rho_i\}$ и $\{\theta_i\}$ - неизвестные параметры, и e - независимые, одинаково нормально распределенные СВ с нулевым средним. Величины y^* представляют собой конечные разности порядка d величин y, а модель обозначается как $ARIMA(p,d,q)$.

Эффективность преобразований AR и MA для устранения автокорреляции остатков продемонстрируем на примере. Нами была оценена зависимость величины реального чистого экспорта RNX для экономики США за 1965 - 1990 гг. от показателей реального ВНП (GNP), его прироста GNP, реального валютного курса ER и его прироста ER. Формула получилась следующей (RNX, GNP - в млрд. долларов; ER - в % к базовому значению):

$$\underset{(34,9)}{RNX = 339,0} - \underset{(0,005)}{0,038 \cdot GNP} - \underset{(0,048)}{0,177 \cdot \Delta GNP} - \underset{(0,24)}{2,15 \cdot ER} + \underset{(0,35)}{2,80 \cdot \Delta ER} \quad (9)$$

(в скобках приведены стандартные ошибки)
$R^2=0,90$; $DW=1,12$.

Значение $DW=1,12$ говорит о наличии некоторой положительной автокорреляции остатков. Использование преобразования $AR(1)$ позволяет существенно улучшить положение; получаем уравнение

$$\underset{(67,2)}{RNX = 323,3} - \underset{(0,014)}{0,035 \cdot GNP} - \underset{(0,035)}{0,155 \cdot \Delta GNP} - \underset{(0,40)}{2,15 \cdot \Delta ER} + \underset{(0,31)}{2,05 \cdot \Delta ER} \quad (10)$$

(в скобках приведены стандартные ошибки)
$R^2 = 0,94$; $DW = 1,71$.

Это уравнение приемлемо по всем параметрам и статистическим характеристикам. Единственное, что имеет смысл сделать в нем, это замена переменных ER и ER на одну переменную $ER(-1)$. Это можно сделать, поскольку абсолютные величины коэффициентов при ER и ER почти одинаковы. В таком случае можно сделать преобразование $(-a \cdot ER + a \cdot \Delta ER) = (-a \cdot ER + a \cdot (ER - ER(-1)) = -a \cdot ER(-1)$, и мы можем использовать это равенство для сокращения числа объясняющих переменных.[1] Включив снова преобразование $AR(1)$ (для которого коэффициент авторегрессии соседних отклонений e_i получился равен $\rho=0,71$, со стандартной ошибкой 0,16), получаем уравнение регрессии:

---

[1] Объединение двух или нескольких переменных осуществляется в уравнении множественной линейной регрессии в том случае, если коэффициенты их близки по абсолютной величине, а новая переменная имеет ясный содержательный смысл. Например, если оценка производственной функции в темповой записи дала результат $y = 1,2 + 0,52 \cdot k - 0,49 \cdot l$ (y, k, l - темпы прироста выпуска, капитала и труда), то можно оценить y как функцию от $(k - l)$. Величина $(k - l)$ - темп прироста капиталовооруженности труда (приближенная оценка).

$$RNX = 318{,}6 - 0{,}035\,GNP - 0{,}155\,\Delta GNP - 2{,}09\,ER(-1) \qquad (11)$$
$$(54{,}9) \quad (0{,}013) \quad (0{,}033) \quad (0{,}28)$$

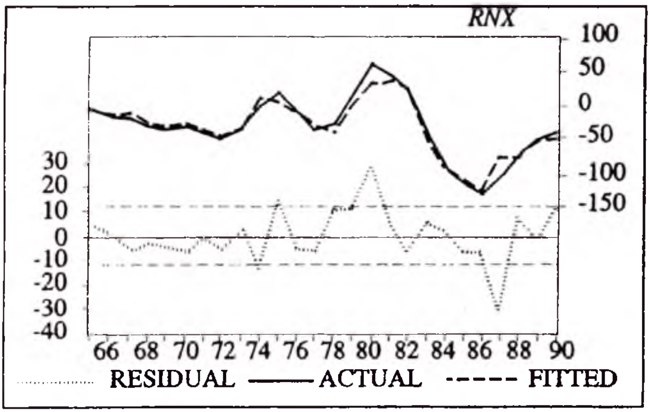
(в скобках приведены стандартные ошибки)

$R^2 = 0{,}94$; $DW = 1{,}79$.

Данное уравнение регрессии приемлемо по всем параметрам и может рассматриваться как конечный результат нашего исследования. Все коэффициенты в нем статистически значимы: даже наименьшая из t-статистик (у коэффициента при переменной GNP) близка к трем. Уравнение регрессии объясняет 94% дисперсии зависимой переменной, а близкая к двум статистика Дарбина-Уотсона не позволяет отвергнуть гипотезу об отсутствии автокорреляции остатков первого порядка. Добавление какой-либо новой объясняющей переменной уже практически не сможет улучшить качества уравнения регрессии.

Качество уравнения показано на рис. 19.1, где фактические и рассчитанные по уравнению регрессии величины очень близки друг к другу, а отклонения от уравнения регрессии выглядят случайными и независимыми друг от друга.

Рисунок 19.1. Действительные (*Actual*) и рассчитанные по уравнению регрессии (*Fitted*) величины RNX; отклонения от линии регрессии (*Residual*). (США, 1965-1990 гг., уравнение: $RNX = 318{,}6 - 0{,}035\,GNP - 0{,}155\,GNP - 2{,}09\,ER(-1)$, с использованием $AR(1)$)



Добавление преобразования $MA(1)$ позволяет повысить величину DW до 1,95, но не дает значимой оценки коэффициента сглаживания и (его t-статистика равна 0,31), поэтому в данном случае включать зто преобразование нецелесообразно.

Преобразования *AR*, *MA* и модель *ARIMA* полезно использовать в тех случаях, когда уже ясен круг объясняющих переменных и общий вид оцениваемой формулы, но в то же время остается существенная автокорреляция остатков. В качестве примера укажем оценивание производственных функций, где объясняющими переменными служат используемые объемы или темпы прироста труда и капитала, а требуемой формулой является, например, производственная функция Кобба-Дугласа или *CES*.

Вопросы к главе 19

1. В каких случаях возникают трудности использования множественной линейной регрессии в моделировании? В чем реальная ситуация может не соответствовать предпосылкам модели?
2. Что такое гомоскедастичность и гетероскедастичность? Каковы результаты использования линейной регрессионной модели в условиях каждой из них? Что можно сказать при этом о несмещенности, состоятельности и эффективности оценок?
3. В чем сущность взвешенного МНК? Какие показатели взвешиваются при его использовании и как могут выбираться весовые коэффициенты?
4. Что такое системы одновременных уравнений в экономическом моделировании? Приведите примеры таких систем.
5. В чем различие структурной и приведенной форм экономических моделей? В чем смысл перехода от структурной к приведенной форме в эконометрике?
6. Как оценить параметры приведенной формы экономической модели? В каких случаях по ним можно рассчитать параметры структурной формы? В чем состоит проблема идентификации в эконометрике и в каких случаях она разрешима?
7. В чем заключается двухшаговый МНК? В каких случаях он применяется?
8. В чем заключается трехшаговый МНК? В каких случаях он применяется?
9. В чем заключается проблема автокорреляции остатков и как она проявляется?
10. Что такое авторегрессионное преобразование и в каких случаях оно применяется?
11. В чем сущность преобразования остатков методом скользящих средних? В каких случаях оно применяется?
12. В чем смысл модели *ARIMA*? Какими могут быть результаты ее применения?

6. **"Искусство менеджмента. Практическое пособие."** (пер. с англ.), 272 стр., переплет.

Книга представляет учебно-практический курс, освещающий все стороны управления коллективом, работающим на фирме, предприятии, в компании. Аргументировано и полно раскрыты функции управления, особенно мотивация, социально-психологические аспекты управления персоналом, процесс принятия решений. Показаны сущность и факторы менеджмента, влияние как внешней, так и внутренней среды на выбор стратегии и методов управления, подбор и обучение кадров, создание психологического климата в коллективе, способствующего успешной реализации поставленной цели.

7. *Д. Бэнгз.* **"Руководство по составлению бизнес-плана"** (пер. с англ.), 256 стр., переплет.

Книга является самым авторитетным и фундаментальным трудом из всех когда-либо изданных на Западе, раскрывающим суть перехода от неформального к формализованному и системному подходу к вопросам планирования бизнеса. Подробно излагаются основы составления бизнес-плана малого предприятия: постановка задачи и выработка путей достижения цели, содержание и форма бизнес-плана, оценка его финансовой стороны. В качестве примеров приводятся реальные бизнес-планы по нескольким проектам, а также образцы оформления вспомогательной документации. **По мнению журнала "Форбс", эта книга — самая лучшая из когда-либо написанных на эту тему.**

8. *Э. Гринолл.* **"Финансы и финансовое планирование для руководителей среднего звена"** (пер. с англ.). 96 стр.

В книге в доступной форме рассматриваются основы финансовой грамотности, процесс финансового планирования и финансовая отчетность на уровне подразделений предприятия. Книга рассчитана на менеджеров среднего звена, предпринимателей и лиц, желающих понять суть и порядок финансового планирования.

9. *Крейнина М.Н.* **"Финансовый менеджмент"**, 320 стр., переплет.

В учебном пособии рассмотрены методы управления финансами предприятия: платежеспособностью, финансовой устойчивостью, кредитоспособностью. Показаны способы управления финансовым состоянием в условиях изменений в спросе на продукцию предприятия. Изложены подходы к финансовому планированию и движению финансовых и денежных потоков с учетом рыночной ситуации и взаимоотношений предприятия с партнерами. Рассмотрены проблемы инвестиционной деятельности с учетом степени рисков. Даны рекомендации выбора финансовых решений в условиях действующей системы налогообложения.

10. *Голубков Е.П.* **"Основы маркетинга"**, 656 стр., переплет.

В основу учебника положен многолетний опыт автора по преподаванию маркетинга студентам экономических учебных заведений, а также слушателям системы повышения квалификации. В книге рассматриваются основные понятия маркетинга и его использование на разных уровнях управления, роль маркетинга в стратегическом планировании. Подробно излагается последовательность функций управления: предплановый анализ, разработка плана, его реализация и контроль выполнения.

Замков О.О., Толстопятенко А.В., Черемных Ю.Н.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ

Учебник

ЛР № 065342 от 12.08.97.

Издательство «Дело и Сервис».
107061, Москва, Б. Черкизовская, 18/1.
Тел. 148-95-62
E-mail: books@aodis. msk.ru.

Подписано в печать с готовых
диапозитивов 02.03.99. Объем 23 п.л.
Формат 60×90/16. Печать офсетная.
Бумага офсетная. Тираж 10 000 экз.
Заказ № 8466.

Государственное унитарное предприятие
Смоленский полиграфический комбинат
Государственного комитета Российской
Федерации по печати.
214020, Смоленск, ул. Смольянинова, 1.

В серии «Учебники Московского государственного университета имени М.В. Ломоносова» в 1998-1999 гг. вышли книги:

- ***«КУРС ЭКОНОМИЧЕСКОЙ ТЕОРИИ»***
  Коллектив авторов под редакцией Сидоровича А.В., объем — 736 стр.
- ***«ГОСУДАРСТВЕННАЯ ЭКОНОМИЧЕСКАЯ ПОЛИТИКА»***
  Коллектив авторов — Албегова И.М., Емцов Р.Г., Холопов А.В., объем — 320 стр.
- ***«МАКРОЭКОНОМИКА», 2-е издание***
  Авторы — Агапова Т.А., Серегина С.Ф., объем — 416 стр.
- ***«МИКРОЭКОНОМИКА», 2-е издание***
  Авторы — Емцов Р.Г., Лукин М.Ю., объем — 320 стр.
- ***«МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ», 2-е издание***
  Авторы — Замков О.О., Толстопятенко А.В., Черемных Ю.Н., объем — 368 стр.
- ***«МЕЖДУНАРОДНАЯ ЭКОНОМИКА»***
  Авторы — Миклашевская Н.А., Холопов А.В., объем — 272 стр.

В издательстве также можно приобрести следующие учебники:

- ***«КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА ХОЗЯЙСТВЕННОЙ ДЕЯТЕЛЬНОСТИ», курс по МВА*** *(пер. с англ.)*
  Автор — Ричард Томас, объем — 432 стр.
- ***«МЕЖДУНАРОДНАЯ СТАТИСТИКА»***
  Авторы — Сиденко А.В., Башкатов Б.И., Матвеева В.М., объем — 272 стр.
- ***«СТАТИСТИКА»***
  Авторы — Сиденко А.В., Матвеева В.М., Башкатов Б.И., Попов Г.Ю., объем — 368 стр.
- ***«МАКРОЭКОНОМИКА»***
  Авторы — Бункина М.К., Семенов В.А., объем — 320 стр.
- ***«КОММЕРЧЕСКИЕ РАСЧЕТЫ от А до Я»***
  Автор — Вебер М., объем — 320 стр.

Справки о приобретении книг по почте и покупке за наличный расчет в г.Москве — по тел. : *(095) 148-81-34, 148-95-62*

Заявки на приобретение направлять по адресу: *107061, г.Москва, а/я 800, ИКЦ «ДИ*