МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ

Малярец Л. М.

ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

Учебное пособие для иностранных студентов

Харьков. Изд. ХНЭУ. 2013

Рецензенты: д.э.н., профессор, заведующий кафедры математики и математических методов экономики Донецкого национального университета Христиановский В.В.; д. ф.-м. н., профессор, заслуженный деятель науки и техники Украины, заведующий кафедры высшей математики Национального университета внутренних дел МВД Украины Яковлев С.В.

Рекомендовано к изданию решением учёного совета Харьковского национального экономического университета.

Протокол № от

Малярец Л.М.
Экономико-математические методы и модели. Учебное пособие для иностранных студентов / Л.М. Малярец. – Харьков: Изд. ХНЭУ, 2013. – с. (рус. яз.)

В пособии изложены все темы нормативных дисциплин «Оптимизационные методы и модели» и «Эконометрика», которые преподаются бакалаврам отрасли знаний 0305 «Экономика и предпринимательство». Изложение материала сопровождается большим количеством примеров, имеющих как учебный, так и исследовательский характер, приведены реальные примеры экономики субъектов хозяйствования различных уровней управления.

Рекомендовано для иностранных студентов, всем студентам, магистрам, аспирантам, изучающим оптимизационным методы и модели, эконометрику.

УДК 519.21 (075.8)
ББК 22.12я73

© Харьковский национальный экономический университет, 2013

© Малярец Л. М., 2013

Введение

Важность моделей известный специалист по эконометрии Э. Маленво так определил: «Все науки используют модели. Каждая выделяет некоторые явления, основные законы которых она и изучает. Исходя из них, она уменьшает сложность действительных ситуаций, упрощает их, чтобы иметь возможность применить два или три закона, которые кажутся нам более важными для изучения проблемы. Изолирование явления и применение к нему абстрактной формализации представляют собой исходную точку любого научного метода» [, с. 49].

Экономико-математические методы являются инструментом изучения, исследования экономических систем разных уровней управления, они составляют фундаментальную основу решения реальных аналитических задач во всех сферах деятельности субъектов хозяйствования.

Все это подчеркивает важность и необходимость преподавания учебной дисциплины «Экономико-математические методы и модели» в формировании компетенций будущих экономистов, аналитиков в процессе подготовки бакалавров и магистров.

Целью создания данного пособия является помочь иностранным студентам, которые учатся в отечественных университетах на экономических направлениях подготовки отрасли знаний «Экономика и предпринимательство», изучить две нормативные дисциплины «Оптимизационные методы и модели» и «Эконометрика». В процессе изучения этих дисциплин студенты приобретут компетенции использования методов математического программирования, эконометрических методов для разработки оптимизационных и эконометрических моделей для решения реальных экономических задач на всех уровнях управления.

Учебное пособие написано согласно действующей министерской программе учебной дисциплины «Экономико-математические методы и модели» для бакалавров отрасли знаний «Экономика и предпринимательство».

Отличительной особенностью данной работы есть изложение материала, содержание и объем которого соответствует лекциям по нормативным дисциплинам «Оптимизационные методы и модели» и «Эконометрика». Теоретический материал сопровождается большим количеством демонстрационных примеров и вопросами для самопроверки. В некоторых темах в качестве примеров представлено решение реальных экономических задач с анализом проблем, которые встречаются при их решении.

Учебное пособие содержит обобщенный материал по 16 темам дисциплины «Оптимизационные методы и модели» и 16 темам дисциплины «Эконометрика». Основными темами учебного пособия есть: концептуальные аспекты математического моделирования в экономике; задача линейного программирования и методы ее решения; симплексный метод решения задач линейного программирования и некоторые его теоретические аспекты; теория двойственности, взаимно двойственные задачи линейного программирования; экономическая интерпретация двойственных неизвестных, двойственный симплекс-метод; анализ линейных моделей экономических оптимизационных задач; транспортная задача; задачи дробно-линейного программирования; целочисленные задачи линейного программирования; методы нелинейного программирования; квадратичное программирование; теория игр, основные методы их решения и анализа; динамическое программирование; парная регрессия и корреляция в эконометрических исследованиях; проверка качества уравнения регрессии; линейные модели множественной регрессии; оценка надежности общей многофакторной линейной модели; мультиколлинеарность, ее последствия и методы устранения; гетероскедастичность и методы ее определения, обобщенный метод наименьших квадратов; автокорреляция остатков модели и методы ее устранения; проблемы интерпретации параметров многофакторной модели; обобщенные схемы регрессионного анализа; системы одновременных уравнений; динамические эконометрические модели; моделирование одномерных временных рядов; моделирование тенденции временного ряда.

Таким образом, приобретенные студентами в результате изучения изложенного материала в пособии знания, навыки, умения формирования оптимизационных и эконометрических задач в экономике, обоснование методов их решения, анализ полученных результатов составляют фундамент компетенций современного экономиста.

Раздел 1. Оптимизационные методы и модели

Тема 1. Концептуальные аспекты математического моделирования в экономике

1.1. Содержание и принципы моделирования

Управленческое решение в экономике будет обоснованным и действенным, если оно разработано с помощью математических методов и моделей. В современных условиях интенсивного развития компьютерной техники и технологии этот факт не требует доказательств.

Известно, что основным методом исследования любых систем является метод моделирования, а именно способ теоретического анализа и практических действий, направленный на разработку и использование моделей. Есть мнение, что развитие любой науки можно трактовать в общем смысле как теоретическое моделирование. Особенное значение приобретают модели в изучении закономерностей массовых процессов, которые недоступны непосредственному наблюдению и не подлежат экспериментированию. Именно экономические системы формируются и развиваются под влиянием бесконечного множества взаимосвязанных факторов и по своей сложности преобладают над техническими, химическими системами.

Модель используется как условный образ, сконструированный для упрощения реального объекта в процессе его изучения. При разработке моделей придерживаются основных методологических принципов: адекватности, динамизма и эвристичности. Принцип адекватности выражает материальное требование к процессу моделирования: необходимость объективного соответствия модели оригиналу как условие объективной истинности знания. В отличие от принципа адекватности, принцип динамизма указывает на изменчивость всех основных элементов процесса моделирования: при изменении целей исследования, изменения объекта, модель должна приспосабливаться к новым условиям и задачам. Принцип замещения утверждает посредническую функцию модели в исследовании. Особенно важно это, если можно провести модельный эксперимент. Принцип эвристичности нацеливает на расширенное воспроизводство знаний. Этими методологическими принципами руководствуются при разработке моделей разных типов.

1.2. Основные типы моделей

Сейчас нет общепринятой классификации моделей. Различают: материальные; знаковые модели, а именно графические и математические; материально-идеальные. Под математической моделью принято понимать совокупность соотношений – уравнений, неравенств, логических условий, операторов и так далее, определяющих характеристики состояние объекта моделирования, параметры функционирования и развития его. Еще математическая модель определяется как гомоморфное отображение в виде упорядоченной совокупности уравнений, неравенств, логических отношений, графиков, условный образ объекта, составленный для упрощения его исследования.

В экономике чаще всего модель экономико-математическая. Она абстрактно описывает экономические системы с помощью формально-математических терминов, логическая структура которых определяется как объективными свойствами предмета описания, так и субъективным целевым фактором исследования, для которого это описание осуществляется. Экономико-математическое моделирование имеет обратное влияние на исследователя, требуя от него четкости формулировки исследовательских задач, строгой логичности в построении гипотез и концепций.

Экономико-математические модели классифицируются:

по способу отображения действительности: аналоговые, концептуальные, структурные, информационные, функциональные;

по признаку целевого предназначения: теоретико-аналитические, прикладные;

по практическому предназначению: балансовые, дескриптивные, имитационные, равновесия, неравновесия, прескриптивные, оптимизационные;

по способу логико-математического описания экономических систем: аналитические, вероятностные, детерминированные, линейные, нелинейные, математико-статистические, матричные, эконометрические;

по временным и пространственным признакам: динамические, статические, точечные, трендовые;

по внутренней структуре модельного описания системы: автономные, глобальные, закрытые, комплекс моделей, макроэкономические, микроэкономические, многосекторые, одно-, многопродуктовые, открытыте;

по области использования: по типу экономических задач, по виду математического метода, применяемого при разработке моделей.

Также предлагается экономико-математические модели разделять на статистические, балансовые и оптимизационные. Статистические модели – это модели, в которых описываются корреляционно-регрессионные зависимости результатов производства от одного или нескольких независимых факторов. Эти модели широко используются для построения производственных функций, а также при анализе экономических систем. Балансовые модели представляют систему балансов производства и распределения продукции и записываются в форме квадратных матриц. Балансовые модели служат для установления пропорций и зависимостей при планировании разных отраслей экономики. Оптимизационные модели представляют систему математических уравнений, линейных или нелинейных, подчиненных определенной целевой функции и служат для отыскания самих лучших (оптимальных) решений конкретной экономической задачи. Эти модели относятся к классу экстремальных задач и описывают условия функционирования экономической системы. Еще оптимизационные модели определяются как экономико-математические модели, охватывающих некоторое количество вариантов производства, распределения и потребления и предназначены для выбора таких значений переменных, характеризующих эти варианты, чтобы был найден лучший из них.

Описание количественных взаимосвязей между признаками осуществляют эконометрические модели. Согласно экономико-математического энцеклопедического словаря эконометрика (эконометрия) научная дисциплина, позволяющая на основе положений экономической теории и результатов экономического измерения придавать конкретные количественные выражения общим (качественным) закономерностям, обусловленным экономической теорией. Эконометрические модели определяются как экономико-математические модели, параметры которых оцениваются с помощью методов математической статистики. Эконометрическая модель выступает в качестве средства анализа и прогнозирования конкретных экономических процессов как на макро- так и на микроэкономическом уровне на основе реальной статистической информации.

Типичными эконометрическими моделями являются производственные функции, выражающие взаимосвязи между затратами и результатами деятельности экономических систем; модели функционирования национальной экономики; типологизация объектов и поведения агентов (стран, регионов, фирм, потребителей); целевые функции потребительского преимущества и функции спроса; модели распределительных отношений в обществе; модели рынка и экономического равновесия; модели интернационализации национальных экономик; модели межрегионального анализа и меж государствами и другие. Опираясь на классификацию задач, решаемых с помощью эконометрики, а именно выделения по конечным прикладным целям, по уровню иерархии и по профилю анализируемой экономической системы, различают прогнозные эконометрические модели, имитационные, модели макро-, мезо- и микроуровня, модели в которых сосредоточено внимание на отдельных проблемах экономических систем.

Процесс построения эконометрической модели объекта в экономике, являющегося экономической системой, имеет свои особенности, выражающиеся в последовательности этапов [28, с. 13 – 14; 21, с. 17 – 23]:

1) анализ характерных признаков экономической системы и процессов, обоснование соответствующей системы показателей и ее структуры взаимосвязей, определение класса моделей, наиболее соответствующих реальному объекту для его описания;

2) формирование совокупности наблюдений – исходных данных для моделирования, являющихся временными, пространственными и пространственно-временными выборками;

3) проверка однородности и точности исходных данных;

4) оценка параметров выбранной формы модели на основе исходных данных;

5) проверка качества вычисленной модели на основе статистических критериев и выводы относительно ее использования для дальнейшего эконометрического исследования;

6) прогнозирование показателей в экономике на основе модели.

В целом экономико-математические модели разрабатываются для достижения трех целей; лучшего понимания объективной реальности; разработки рационального варианта действий, а также для выбора оптимальных решений в практической деятельности.

1.3. Этапы экономико-математического моделирования

Если в моделировании руководствоваться правильной методикой, то разработанные модели будут адекватные реальному объекту, допустимые с точки зрения вычислительных процедур, а сам процесс моделирования – эффективным и обоснованным. В таком случае моделирование достигает цели. Методу моделирования можно рассматривать как последовательную логику взаимосвязанных этапов: определение цели моделирования, анализ сформулированной проблемы и разработка концептуальной модели; построение математической модели и ее анализ; подготовка исходных данных; численное решение модели; анализ числовых результатов на непротиворечивость с концептуальной моделью; использование модели для обоснования управленческого решения [23; 33].

Цель моделирования следует формулировать, исходя из сущности проблемы исследования. Поэтому изначально формируются гипотезы, объясняющие поведение и развитие моделируемого объекта, а также прогнозируют цели, которые будут изменять во времени.

На этапе анализа сформулированной проблемы и разработки концептуальной модели предусматривается выделение пределов экономической системы, ее структуризация. Под структурой экономической системы понимают ее статическое представление в разрезе материальных и нематериальных элементов, обеспечивающих ее форму и организованность. Процесс логического разделения большой проблемы на отдельные элементы предусматривает получение оптимального управленческого решения, принимаемого на основе экономико-математической модели. Предназначение концептуальных моделей – содержательно представлять существенные свойства объекта и главные связи между этими свойствами.

На этапе построения математической модели осуществляется формализация концептуальной модели. Математическая формализация обозначает, что отработаны конкретные правила действий, концептуальные положения, адекватные целям исследования и принятой системе гипотез, осуществляется глубинная связь между математическим инструментом, предметом исследования и исследователем. Этап математического анализа модели связан с тем, что математическими приемами исследования выявляются общие свойства модели и ее решений, при этом важным моментом является доказательство существования решения сформулированной задачи.

На этапе подготовки исходных данных предусматривается количественное измерение признаков объекта, которые являются его основными свойствами и отображение величин в системе показателей. Информационной моделью в экономико-математическом моделирования считают иерархическую систему показателей, отражающих признаки объекта.

На этапе численного решения модели используют существующие программные средства, прикладные пакеты, разрабатываются специальные вычислительные программы для реализации модели. Вычисления могут иметь многовариантный характер для изучения поведения модели в разных условиях и ограничениях.

Анализ числовых результатов и их использование содержат проверку правильности и полноты результатов моделирования и использования в практической деятельности, а также для совершенствования самой модели. На этом этапе следует выполнить верификацию (проверку правильности структуры, логики модели) и валидацию модели (проверку соответствия данных, полученных на основе модели, реальному процессу).

Этап использования результатов моделирования для принятия управленческого решения состоит из качественного анализа полученных результатов, которые не только представляются формулами, но и для наглядности изображаются в виде графиков, таблиц, схем.

Отдельно следует указать, что каждый этап моделирования должен сопровождаться оценкой полученных результатов для своевременного устранения выявленных ошибок. Самыми типичными ошибками является включения в модель несущественных для данной задачи переменных, не включения в модель существенных переменных, не достаточно точная оценка параметров модели, недостатки в структуре модели, что приводит к неправильной спецификации модели. Процесс моделирования имеет циклический характер и, начиная моделирование объекта с разработки простой модели переходят к разработки сложных моделей, совершенствование их с помощью учета новых условий и уточнения математических зависимостей.

В математическом моделировании различают такие виды контроля; контроль характера зависимостей, контроль экстремальных ситуаций, контроль предельных условий, контроль математической замкнутости. С помощью контроля характера зависимостей проверяется направление и скорость изменений одних признаков при и изменении других. Контроль экстремальных ситуаций направлений на исследование вида, принимающего как изначальный и промежуточные соотношения, а также выводы, следующие в процессе реализации модели, если параметры или их комбинации приближаются к крайним допустимым для них значениям, чаще всего к нулю или к бесконечности. В таких экстремальных ситуациях задача часто упрощается или вырождается, так что соотношения приобретают более научное содержание и могут быть проще проверены, а конечные выводы могут быть продублированы каким-нибудь другим методом. Если в процессе исследования математической модели должна быть построена некоторая функция, то обычно необходимо, чтобы на пределе области своего определения удовлетворяла указанным предельным условиям, вытекающим из содержания задачи. Контроль математической замкнутости состоит в проверке того, что выписанные соотношения предоставляют возможности, причем однозначно, решить поставленную задачу.

В целом процесс экономико-математического моделирования является преобразование моделей, изображающего следующей цепью: когнитивная модель $\rightarrow$ содержательная модель (описательная, объясняющая, предполагаемая) $\rightarrow$ концептуальная модель (логико-семантическая, структурно-функциональная, причинно-следственная) $\rightarrow$ формализованная модель (математическая и информационная). Когнитивной принято называть сформированный в голове исследователя некоторый мысленный образ. Представление когнитивной модели обычным языком является содержательной моделью. Концептуальной моделью называют содержательную модель, при формировании которой используют понятия и представления предметной области знаний, к которой принадлежит объект моделирования. Формализованная модель является представление концептуальной модели с помощью одного или нескольких формальных языков. Математическое моделирование является идеальным научным знаковым формализованным моделированием, при котором описание объекта осуществляется на языке математики, а исследование модели выполняется с помощью математических методов.

Вопросы для самопроверки:

1) Какие основные методологические принципы моделирования?
2) Какая основная суть каждого методологического принципа моделирования?
3) Дать определение математической модели.
4) Какие основные типы моделей?
5) Как классифицируются экономико-математические модели?
6) Какие особенности оптимизационных моделей?
7) Какие особенности эконометрических моделей?
8) Из каких этапов состоит построение эконометрических моделей?
9) Из каких этапов состоит технология экономико-математического моделировании?
10) Какая модель называется когнитивной?
11) Какая модель называется содержательной?
12) Какая модель называется концептуальной?
13) Какая модель называется формализованной?

Тема 2. Оптимизационные экономико-математические модели

2.1. Классификация оптимизационных задач

2.2. Основы классической теории оптимизации

2.3. Общая постановка задачи оптимизации

2.4. Классическая задача условной оптимизации. Формулирование задачи

2.5. Метод множителей Лагранжа

Большую группу математических методов в экономике образуют оптимизационные методы, поскольку перед менеджерами, экономистами, работниками системы управления, инженерами разного уровня возникают проблемы принятия решения, требующие оптимизации результатов разных видов деятельности с учетом имеющихся ресурсов. Существует специальная теория принятия решения, имеющая экономические, психологические, политические, социальные, финансовые и другие аспекты, учитывая которые следует искать самые лучшие решения. Алгоритмизация процесса разработки модели предусматривает использование разных математических методов для поиска самого лучшего решения. Эти методы делятся на методы линейной и нелинейной оптимизации.

Типичными оптимизационными задачами являются задачи оптимального планирования, в которых выделяют переменные и параметры (количество покупаемых продуктов, количество произведенной продукции, количество перевозимого груза), цель, которую желают достичь (функция цели) и которую следует оптимизировать (минимизировать затраты на потребление, максимизировать прибыль предприятия, минимизировать стоимость перевозок) и ограничения то есть у условия, ограничивающие возможности достижения желаемой цели (в рационе должны быть определены компоненты, ограничены ресурсы предприятия, количество перевозимого товара). Целевая функция имеет смысл ожидаемой «ценности» или «полезности». Задача оптимального планирования также называется оптимизационной или экстремальной задачей.

В задачах оптимизации могут быть выделены характеристики объекта (объектов), которыми можно или надо варьировать для достижения цели. Такие характеристики называются управляемыми переменными или управляемыми параметрами. Набор значений управляемых переменных в задачах оптимизации называются решением. Значения управляемых переменных могут быть ограниченными. Решение, удовлетворяющее выдвинутым ограничениям называют допустимым решением. Оптимальным называется допустимое решение, которое по некоторым причинам преобладает над другими решениями, например, решение, при котором целевая функция экстремальная. Существуют также неуправляемые переменные, то есть изменение их значений не зависит от управляющего субъекта.

Задачи оптимизации в экономике объединяются в разделе «Исследование операций», который предусматривает использование математических методов для моделирования и анализа ситуаций в экономике. Среди этих методов отдельно выделяют линейное и нелинейное программирование, называя математическим программированием, являющегося одним из основных методов принятия производственно-экономических решений.

2.1. Классификация оптимизационных задач

Любая модель в оптимизационной задачи включает искомые переменные, налагаемые на них ограничения и формулировку цели. Цель определяет целевую функцию, задаваемую на множестве допустимых решений D. Множество D выражает меру осуществления цели: если D пусто, то решения не существует; если D содержит одну точку, то эта точка является единственным допустимым решением задачи и такая задача не представляет интереса для исследования; если D содержит более чем одно решение, тогда задача оптимизации заключается в нахождении оптимального решения на множестве допустимых решений. При этом, если D конечно, то оптимальное решение может быть найдено в результате простого перебора всех точек D и вычисления в них функции цели. Если D счетное или D есть континуумом, то оптимальное решение следует искать в бесконечном множестве допустимых решений.

Математическая модель задачи оптимизации предусматривает, что все переменные, параметры, ограничения и целевая функция модели количественно измеримы. Если переменные $X = (x_1, x_2, ..., x_n)$ являются управляемыми переменными, а $Z = (z_1, z_2, ..., z_k)$ - не управляемыми параметрами и условие функционирования исследуемой системы определяются m ограничениями, то математическую модель записывают так:

найти точку $Y = (y_1, y_2, ..., y_n)$, в которой достигается экстремум, минимум или максимум, целевой функции $f(X, Z)$:

$$f(Y, Z) = extrf(X, Z),$$

при ограничениях

$$\varphi_i(X, Z) \leq (\geq, =) b_i, \; i = \overline{1,m},$$

$$x_j \geq 0, \;\; j = \overline{1,n}.$$

Таким образом, задача на условный экстремум обычно имеет смысл, если $m < n$.

Оптимизационные задачи по виду целевой функции и по виду ограничений классифицируются на группы. Если функция цели и система ограничений являются линейными, то говорят о линейном программировании, в противном случае возникает задача нелинейного программирования. В случае квадратичной функции цели и линейной системы ограничений задачу оптимизации называют задачей квадратичного программирования. Если функцию цели можно представить в виде суммы таких функций, что каждая из них зависит только от одной переменной, то рассматривают задачу сепарабельного программирования. Если управляемые переменные принципиально могут быть только целыми, то такая задача оптимизации называется целочисленной.

Если функция цели является выпуклой функцией, то такая задача оптимизации называется задачею выпуклого программирования. Если функции φ_i, определяющие ограничения $\varphi_i(X, Z) \le (\ge, =)\, b_i$ являются выпуклыми вверх (выпуклыми вниз) функциями, то они порождают выпуклое множество допустимых решений.

По информационным свойствам оптимизационные задачи делятся на статические и динамические. Если субъект в процессе принятия решения не меняет свое информационное состояние, то принятие решения рассматривается как мгновенный факт и такие задачи называются статическими. Если субъект в процессе принятия решения меняет свое информационное состояние, то решение принимается поэтапно и задача называется динамической.

Кроме приведенных моделей, к оптимизационным моделям иногда относят имитационные модели и эвристические. Имитационные модели позволяют имитировать поведение очень сложных систем, для которых не возможно построить математические модели и получить решение. Здесь возникают сложности связанные с разработкой эксперимента, проверкой статистической значимости его результатов, а также сам процесс оптимизации вызывает усложнения. Если имеющаяся задача оптимизации настолько сложна, но существует предположение, что решение есть, то используют разные эвристические методы, которые основаны на интуиции исследователя или на эмпирических данных. Следует указать, что точность модели также зависит от объема и состава имеющихся исходных данных.

2.2. Основы классической теории оптимизации

Пусть функция $f(X)$ определена в области O n-мерного евклидова пространства $\Re^n$, $X = (x_1, x_2, ..., x_n) \in O \subseteq \Re^n$ и принимает значения на некотором интервале O_f действительной оси, $f = (X) \in O_f \subseteq \Re$. Точка

$Y = (y_1, y_2, ..., y_n) \in O$ называется точкой локального минимума или локального максимума функции $f(X)$, если существует такое $\varepsilon > 0$, что неравенство

$$f(Y) \leq f(Y + \delta X) \text{ или } f(Y) \geq f(Y + \delta X) \qquad (2.1)$$

Выполняется для всех $\delta X = (\delta x_1, \delta x_2, ..., \delta x_n)$, таких, что $0 < |\delta x| < \varepsilon$. Значения функции $f(Y)$ в этой точке называется локальным минимумом или локальным максимумом. Множество точек $U_\varepsilon(Y) = \{Y + \delta X\}$ образует открытую сферу радиуса ε с центром в точке $У$ или ε-окрестность. Если неравенство (2.1) выполняется строго,

$$f(Y) < f(Y + \delta X) \text{ или } f(Y) > f(Y + \delta X), \qquad (2.2)$$

то такой минимум (максимум) называется строгим, в противном случае минимум (максимум) – нестрогий. Поиск точек минимума или максимума это поиск точек экстремума.

Точка $Y \in O$ называется точкой глобального минимума (максимума) функции $f(X)$, если неравенство (2.1) или (2.2) выполняется для всех точек δX области определения O ($\delta X \in O$)

$$f(Y) \leq f(X) \text{ или } f(Y) \geq f(X).$$

Известно, что точка глобального экстремума есть точкой локального экстремума, а обратное утверждение неверно.

Если точка $Y \in O$ есть точкою, в которой функция $f(X)$ достигает глобального минимума или максимума, то из определения экстремума вытекает, что число $f^* = f(Y)$ есть соответственно точной нижней границей или точкой верхней границей множества значений функции $f(X)$:

$$f^* = \inf_{X \in O} f(X) = f(Y) \text{ или } f^* = \sup_{X \in O} f(X) = f(Y).$$

Обратное может быть неверно, если функция $f(X)$ не имеет экстремума, то

$$f^* = \inf_{X \in O} f(X) < f(X) \text{ или } f(X) = \sup_{X \in O} f(X) = f^*.$$

От локального или глобального минимума (максимума) функции $f(X)$ всегда можно перейти к локальному или глобальному максимуму (минимуму) функции $f(X)$ при помощи преобразования

$$F(X) = -f(X).$$

Функции, имеющие единственный минимум (максимум), называются унимодальными.

2.3. Общая постановка задачи оптимизации

В общем виде задача оптимизации, или задача определения экстремума ставится следующим образом.

Пусть заданы:

функция $f(X)$, определенная на множестве $O \subseteq \Re^n$;

множество $D \subseteq \Re^n$.

Найти точку $Y = (y_1, y_2, ..., y_n) \in D$, в которой функция $f(X)$ достигает экстремального (минимального или максимального) значения, т.е.

$$f(Y) = extr\ f(X) \text{ и } Y \in D.$$

Функция $f(X)$ называется целевой функцией, переменные X – управляемые переменными, D – допустимым множеством, а любой набор значений $У$ управляемых переменных, принадлежащий D ($Y \in D$)– допустимым решением задачи оптимизации.

Точка $У$, в которой $f(X)$ достигает своего экстремума, должна принадлежать пересечению области определения O функции $f(X)$ и допустимого множества D ($Y \in O \cap D$). Если множества O и D совпадают со всем пространством $\Re^n$ $(O = D = \Re^n)$, то такая задача называется задачей на безусловный экстремум. Если хотя бы одно из множеств O или D является собственным подмножеством пространства $\Re^n$ $(O \in \Re^m, D \in \Re^n)$ или множества O и D пересекаются $(O \cap D \neq \varnothing)$, то такая задача называется задачей на условный экстремум, в противном случае $(O \cap D = \varnothing)$ точка экстремума $У$ не существует. Если множества O и D пересекаются в одной точке $У$, то эта точка $У$ является единственным допустимым решением и обсуждать проблему поиска экстремума бессмысленно.

Обычно в задаче условного экстремума задается не само допустимое множество решений D, а система соотношений, его определяющая,

$$\varphi_i(x_1, x_2, ..., x_n) \leq (=, \geq)0, i = \overline{1,m},$$

$$\text{то есть } D = \{\varphi_i(x_1, x_2, ..., x_n) \leq (=, \geq)0, i = \overline{1,m}\} \subseteq \Re^n,$$

или множество D может одновременно задаваться как в явном виде, т.е. допустимое решение X должно принадлежать некоторой области $P \in \Re$, так и системой ограничений.

Таким образом, задача на безусловный экстремум целевой функции $f(X)$ заключается в нахождении точки $Y = (y_1, y_2, ..., y_n) \in \Re^n$, в которой достигается экстремум $f(X)$:

$$f(Y) = extr\ f(X). \tag{2.3}$$

Задача на условный экстремум целевой функции $f(X)$ заключается в нахождении точки $Y = (y_1, y_2, ..., y_n) \in P \subseteq \Re^n$, в которой достигается экстремум $f(X)$:

$$f(Y) = extr\ f(X), \tag{2.4}$$

при условии, то точка $У$ удовлетворяет ограничениям

$$\varphi_i(x_1, x_2, ..., x_n) \leq (=, \geq) 0, i = \overline{1, m}, \tag{2.5}$$

где $\varphi_i(X)$ заданы на некотором множестве $O' \in \Re^m$, т.е.

$$Y \in O \bigcap O' \bigcap D$$

$$D = \{X : X \in P\ i\ \varphi_i(X) \leq (=, \geq) 0, i = \overline{1, m}\} \subseteq \Re^n.$$

Здесь $O \bigcap O'$ - область определения функций $f(X)$ і $\varphi_i(X)$.

Как правило, в задаче условной оптимизации $D \subseteq O \bigcap O'$ и поэтому область определения функций $f(X)$ и $\varphi_i(X)$ явно не указывают, а рассматривают только множество P, всегда подразумевая, что функция цели $f(X)$ и функции $\varphi_i(X)$, определяющие ограничения, по крайней мере, заданы на P. Естественно, что при необходимости ограничения, обусловленные множествами O и O', включаются в P.

В решении реальных экономических проблем задачи нахождения безусловного экстремума, как правило, встречаются чрезвычайно редко.

Следует отметить, что если $У$ – точка безусловного локального экстремума функции $f(X)$ находится внутри множества допустимых значений D , то ограничения выполняются автоматически, если точка $У$ - за пределами множества допустимых значений D, то точка условного локального экстремума $f(X)$ находится на границе D, т.е. ограничения-неравенства превращаются в ограничения-равенства.

В задаче только с ограничениями-равенствами количество ограничений m должно быть меньше числа переменных n, $m<n$, так как при $m>n$ по крайней мере (m-n) ограничений оказываются избыточными либо система ограничений (2.5) либо система ограничений (2.5) переопределена и, следовательно, не имеет допустимого решения. При $m=n$ допустимое решение, скорее всего, будет единственным, и такая задача оптимизации также не представляет интереса. Задача поиска оптимального решения имеет смысл, когда $m<n$.

Если $m<n$, то всегда с n переменных $X = (x_1, x_2, ..., x_n)$ можно выбрать n-m переменных, которые называются независимыми (свободными) переменными; остальные переменные называются зависимыми переменными.

2.4. Классическая задача условной оптимизации. Формулирование задачи

Рассмотрим задачу условной оптимизации (2.4) – (2.5), которая заключается в нахождении точки $Y = (y_1, y_2, ..., y_n) \in P \subseteq \Re^n$, в которой достигается экстремум $f(X)$:

$$f(Y) = extr\ f(X), \tag{2.4}$$

при условии, то точка $У$ удовлетворяет ограничениям

$$\varphi_i(x_1, x_2, ..., x_n) \leq (=, \geq)\, 0, i = \overline{1,m}. \tag{2.5}$$

Сформулируем необходимые и достаточные условия существования экстремума в задаче на условный экстремум.

Пусть вектор $Z \in D \subseteq \Re^n$ задает возможное направление на множестве допустимых значений D, если при всех достаточно малых $\alpha > 0$ вектор $Y + \alpha Z \in D$. Все возможные направления на множестве допустимых значений D функции $f(X)$ в точке $У$ образуют множество возможных направлений $v(Y, f) \subseteq \Re^n$.

Теорема. Если точка $У$ является точкой локального минимума (максимума) задачи (2.4) – (2.5), то

$$w(Y, f) \cap v(Y, f) \neq \varnothing,$$

где $w(Y, f)$ – множество направлений убывания (возрастания) $f(X)$ в точке $У$.

Дана теорема показывает, что невозможно смещение из точки локального экстремума $У$, которое приводит к уменьшению (увеличению) целевой функции и при этом не выходит за пределы допустимой области. Автор книги

«Введение в теорию и методы оптимизации для экономистов» Фролькис В.А. [№] считает, что данная теорема тривиальна и ее следует дополнить теоремой Вейерштрасса, а именно:

Теорема Вейерштрасса. Пусть $D \subseteq \Re^n$ – замкнутое ограниченное множество и $f(X)$ – непрерывная функция, определенная на *D*. Тогда существует точка $Y \in D$ глобального минимума (максимума) функции $f(X)$, $f(Y) = min(max) f(X)$.

Приведенные две теоремы не указывают алгоритм нахождения экстремальной точки. Допустим, функция $f(X)$ определена и непрерывна в ограниченной замкнутой области *D*. В этой области (согласно теореме Вейерштрасса) найдется точка, в которой функция будет экстремальна. В действительности свое экстремальное значение функция $f(X)$ может достигнуть как во внутренней точке области *D*, так и на ее границе. Если дополнительно функция $f(X)$ имеет конечные частные производные первого и второго порядка, то для получения необходимых и достаточных условий существования локального экстремума целевой функции $f(X)$ во внутренних точках *D* можно использовать теоремы, определяющие необходимые и достаточные условия существования локального безусловного экстремума, но они не диагностируют экстремальные точки, лежащие на границе *D*. Поэтому для того, чтобы найти глобальный экстремум $f(X)$ в области *D,* необходимо найти все внутренние точки локального экстремума функции $f(X)$, вычислить в них значения $f(X)$ и сравнить со значениями функции $f(X)$ на границе области. Наименьшее (наибольшее) из них и будет искомым глобальным значением. Если внутренних стационарных точек нет, то экстремальное значение может достигаться только на границе области. Во многих задачах условной оптимизации экстремум достигается именно на границе области *D,* поэтому для таких задач результаты теорем классического анализа неприемлемы.

Для того чтобы не вычислять значения целевой функции вдоль границы области допустимых значений *D*, а сформулировать необходимые и достаточные условия экстремума как для внутренних точек области *D,* так и для точек на ее границе, рассмотрим частный случай задачи (2.4) – (2.5) в виде, который называется классическая задача оптимизации.

Найти точку локального экстремума $Y = (y_1, y_2, ..., y_n)$ целевой функции $f(X)$,

$$f(Y) = extr\ f(X), \tag{2.6}$$

При условии, что точка $У$ удовлетворяет ограничениям-равенствам

$$\varphi_i(X) = 0, i = \overline{1,m},\ m < n, \tag{2.7}$$

т.е. $Y \in D$, где допустимое множество определяется как

$$D = \{X:\ \varphi_i(X) = 0, i = \overline{1,m}\} \subseteq \Re^n \tag{2.8}$$

и $D \subseteq O \bigcap O'$.

В теории условной оптимизации существует несколько методов, позволяющих свести задачу условного экстремума к задаче безусловного экстремума и благодаря такому приему использовать необходимые и достаточные условия экстремума. К таким наиболее распространенным методам относится метод множителей Лагранжа и метод Якоби.

2.5. Метод множителей Лагранжа

Метод множителей Лагранжа позволяет свести классическую задачу на условный экстремум к задаче на безусловный экстремум. Идея метода Лагранже состоит в построении вспомогательной функции, называемой функцией Лагранжа, точка безусловного локального экстремума $У$ совпадает с точкой условного локального экстремума функции задачи (2.6) – (2.8).

Такими свойствами обладает функция вида

$$L(\Lambda, \lambda_0, X) = \lambda_0 f(X) + \sum_{i=1}^{m} \lambda_i \varphi_i(X), \tag{2.9}$$

параметры λ_0 i $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_m) \in \Re^m$ которой называются множителями Лагранжа. Функцию Лагранжа можно записать в виде:

$$L(\Lambda, \lambda_0, X) = \lambda_0 f(X) - \sum_{i=1}^{m} \lambda_i \varphi_i(X).$$

Если функции $f(X)$ и $\varphi_i(X)$ непрерывны и имеют частные производные первого порядка, то частные производные функции $L(\Lambda, \lambda_0, X)$ по переменным x_j имеют вид:

$$\frac{\partial L(\Lambda, \lambda_0, X)}{\partial x_j} = \lambda_0 \frac{\partial f(X)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial \varphi_i(X)}{\partial x_j}. \tag{2.10}$$

Составленный из них вектор обозначается как:

$$L'_x(\Lambda, \lambda_0, X) = \lambda_0 f'(X) + \sum_{i=1}^{m} \lambda_i \varphi'_i(X).$$

Частные производные функции $L(\Lambda, \lambda_0, X)$ по переменным $\lambda_i, i = \overline{1,m}$ всегда существуют и равны:

$$\frac{\partial L(\Lambda, \lambda_0, X)}{\partial \lambda_i} = \varphi_i(X). \qquad (2.11)$$

Необходимые условия экстремума первого порядка в задаче безусловной оптимизации функции Лагранжа формулируются в виде следующей теоремы.

Теорема (Правило множителей Лагранжа). Если точка $Y \in D \subseteq \Re^n$ – точка локального оптимума задачи на условный экстремум (2.6) – (2.8):

$$f(Y) = extr\ f(X) \text{ и } \varphi_i(X) = 0, i = \overline{1,m}$$

в окрестности некоторой функции $f(X)$ и $\varphi_i(X)$ непрерывно дифференцируемы, и в точке *У* ранг матрицы Якоби равен *m*, то существуют такие неравные одновременно нулю вектор Λ' и параметр λ'_0, что точка $(\Lambda', \lambda'_0, Y)$ является стационарной точкой задачи на безусловный экстремум функции Лагранжа (2.9), соответствующей задаче (2.6) – (2.8):

$$grad\ L(\Lambda', 1, Y) = 0.$$

Учитывая вид производных (2.10) – (2.11) стационарную точку *У* функции Лагранжа можно определить как точку, удовлетворяющую условиям

$$L'_x(\Lambda, \lambda_0, X) = 0, \varphi_i(X) = 0, i = \overline{1,m}.$$

В отличие от задачи на безусловный экстремум стационарная точка функции Лагранжа может находиться на границе области *D*.

Приклад. Найти точки условного экстремума функции $f(X) = xyz + 8x^2 + 12y^2 + 25z^2$ при ограничениях

$$x + y + 7z = 12,$$
$$6x + 3y - 4z = 25.$$

Функция Лагранжа рассматриваемой задачи:

$$L(\Lambda, X) = xyz + 8x^2 + 12y^2 + 25z^2 + \lambda_1(x + y + 7z - 12) + \lambda_2(6x + 3y + 4z - 25).$$

Для того, чтобы точка (Λ', Y) была стационарной точкою $L(\Lambda, X)$, должны выполняться необходимые условия экстремума:

$$\frac{\partial L(\Lambda, X)}{\partial x} = 16x + xy + \lambda_1 + 6\lambda_2 = 0,$$

$$\frac{\partial L(\Lambda, X)}{\partial y} = xy + 24y + \lambda_1 + 3\lambda_2 = 0,$$

$$\frac{\partial L(\Lambda, X)}{\partial z} = xy + 50z + 7\lambda_1 + 4\lambda_2 = 0,$$

$$\varphi_1(X) = x + y + 7z - 12 = 0, \quad \varphi_2(X) = 6x + 3y + 4z - 25 = 0.$$

Решением системы являются две стационарные точки (Λ^1, Y^1) і (Λ^2, Y^2) с координатами:

$$Y^1 = (2{,}917;\ 0{,}95;\ 1{,}162), \quad \Lambda^1 = (-4{,}581;\ -7{,}2),$$

$$Y^2 = (13036;\ -245{,}198;\ 20{,}595), \quad \Lambda^2 = (3872;\ -105{,}217),$$

$$f(Y^1) = 115{,}877; \quad f(Y^2) = 263500\ .$$

С помощью частных производных второго порядка доказывается, что точка Y^1 является точкой минимума функции $f(X)$, а точка Y^2 - точка максимума.

Рассмотрим экономическую интерпретацию множителей Лагранжа. Для задачи на условный экстремум:

найти точку локального экстремума $Y = (y_1, y_2, ..., y_n)$ целевой функции $f(X)$,

$$f(Y) = extr\ f(X),$$

При условии, что управляемые переменные $X = (x_1, x_2, ..., x_n)$ удовлетворяют ограничениям-равенствам

$$\varphi_i(X) = b_i, i = \overline{1,m}.$$

Если целевую функцию $f(Y)$ в точке экстремума *У* интерпретировать как оптимальный доход (или стоимость), а постоянные b_k $-$ как объемы каких-либо ресурсов, то множители Лагранжа λ_k показывают, как изменится оптимальный доход (или стоимость) при увеличении количества *k*-го ресурса на единицу. Другими словами, множители Лагранжа указывают, как изменится оптимальное значение целевой функции $f(Y)$ при изменении параметра b_k в правой части ограничений на единицу, т.е. определяют ценность *k*-го ресурса.

Вопросы для самопроверки:

1) Привести классификацию оптимизационных задач.

2) Привести постановку общей оптимизационной модели.
3) Как точка называется точкой экстремума?
4) Как определяется допустимое множество в оптимизационной модели?
5) Как формулируется задача на безусловный экстремум?
6) Как формулируется задача на условный экстремум?
7) Какая точка называется точкой локального минимума (максимума)?
8) Какие методы позволяют привести задачу условного экстремума к задаче безусловного экстремума?
9) Как строится функция Лагранжа?
10) Сформулировать правило множителей Лагранжа.
11) Какая экономическая интерпретация множителей Лагранжа?

Тема 3. Задача линейного программирования и методы ее решения

3.1. Постановка задачи линейного программирования (ЛП)
3.2. Каноническая форма задачи линейного программирования
3.3. Графическое решение задач математического программирования
3.4. Свойства возможных решений задачи линейного программирования

3.1. Постановка задачи линейного программирования (ЗЛП)

Большинство задач в экономике являются оптимизационными. Одной из самых распространенных задач математического программирования есть задача оптимизации производства, например, разработка плана выпуска максимального объема продукции при ограниченных сырьевых ресурсах или в пределах конкретного плана производства определение минимально необходимого объема производственных затрат. К этой группе принадлежит задача о диете, где необходимо учитывать калорийность и потребительскую ценность продуктов, а также их стоимость; задачи об определении оптимального состава смесей при кормежке животных. К ЗЛП принадлежит задача о раскрое материалов, суть которой состоит в разработке таких технологически допустимых планов раскроя, при которых получается необходимый комплект заготовок, а отходы сводятся к минимуму. Другим наиболее распространенным классом задач является задача о минимизации транспортных затрат, когда речь идет о рациональной перевозке некоторого однородного продукта от производителей к потребителям.

Рассмотрим типовые задачи линейного программирования.

Задача об оптимальном использовании ресурсов

Пусть для производства некоторой продукции A_1, A_2, A_3 используются ресурсы B_1, B_2, B_3, B_4 (сырье, полуфабрикаты, оборудование, финансы), запасы которых составляют b_1, b_2, b_3, b_4. Известны нормы затрат a_{ij} каждого вида ресурса B_i на изготовление единицы продукции A_j, а также прибыль c_j от производства и реализации каждого вида продукции (табл. 3.1).

Таблица 3.1

Условие задачи

Ресурсы	Продукция			Запасы ресурсов
	A_1	A_2	A_2	
B_1	a_{11}	a_{12}	a_{13}	b_1
B_2	a_{21}	a_{22}	a_{23}	b_2
B_3	a_{31}	a_{32}	a_{33}	b_3
B_4	a_{41}	a_{42}	a_{43}	b_4
Прибыль	c_1	c_2	c_3	

Необходимо найти план производства, т.е. необходимо изготовить такое количество x_1, x_2, x_3 продукции каждого вида, чтобы максимизировать общую прибыль. Поскольку прибыль от производства единицы A_1 равна c_1, то прибыль от планового количества x_1 буде равняться произведению $c_1 x_1$; общая прибыль будет равна сумме прибылей от производства всех видов продукции:

$$z = c_1 x_1 + c_2 x_2 + c_3 x_3 \to max .$$

Затраты ресурса B_1 на производства запланированных объемов продукции x_1, x_2, x_3 будут равны произведениям $a_{11} x_1$, $a_{12} x_2$, $a_{13} x_3$. Необходимо учесть, что общие затраты этого ресурса $a_{11} x_1 + a_{12} x_2 + a_{13} x_3$ не могут превышать его запасы b_1, т.е. : $a_{11} x_1 + a_{12} x_2 + a_{13} x_3 \le b_1$. Запишем аналогичные условия по всем видам ресурсов и получим систему неравенств:

$$\begin{cases} a_{11} x_1 + a_{12} x_2 + a_{13} x_3 \le b_1 ; \\ a_{21} x_1 + a_{22} x_2 + a_{23} x_3 \le b_2 ; \\ a_{31} x_1 + a_{32} x_2 + a_{33} x_3 \le b_3 ; \\ a_{41} x_1 + a_{42} x_2 + a_{43} x_3 \le b_4 . \end{cases}$$

Додадим еще естественное требование неотрицательности неизвестных

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

Математическая модель этой задачи имеет вид:

Необходимо найти максимум линейной функции цели

$$z = \sum_{j=1}^{n} c_j x_j \to max$$

при имеющейся линейной системе ограничений-неравенств:

$$\begin{cases} \sum\limits_{j=1}^{n} a_{ij} x_j \le b_i \\ (i = \overline{1, m}) \end{cases},$$

где все $x_j \ge 0$.

Где n – количество видов продукции, m – количество видов ресурсов. Это есть задача линейного программирования (ЗЛП).

Модель, где требуется найти максимум линейной функции цели при линейных ограничениях типа $\le$ является 1-й стандартной формой ЗЛП. Если продукция измеряется в штуках, то в модель еще требуется включить условие, чтобы все неизвестные были целыми.

Задача составления самого дешевого рациона (кормовой смеси) или задача о смесях (задача о диете)

С имеющихся в продаже продуктов необходимо составить такой рацион питания, чтобы его общая стоимость была наименьшей, при этом должны были обеспечены все потребности организма в питательных веществах (белки, жиры, углеводы). В табл. 3.2 приведены данные о содержании питательных веществ B_1, B_2, B_3 у каждом виде продуктов A_1, A_2, A_3, A_4, стоимости продуктов и минимальные потребности организма в питательных веществах.

Таблица 3.2

Условие задачи

Продукты	Питательные вещества			Стоимость продуктов
	B_1	B_2	B_3	
A_1	a_{11}	a_{12}	a_{13}	c_1
A_2	a_{21}	a_{22}	a_{23}	c_2
A_3	a_{31}	a_{32}	a_{33}	c_3
A_4	a_{41}	a_{42}	a_{43}	c_3
Ежедневный спрос	b_1	b_2	b_3	

Необходимо определить, с какого количества каждого вида продуктов должен состоять наиболее дешевый рацион. Обозначим эти неизвестные количества (состав рациона) через x_1, x_2, x_3, x_4. Стоимость единицы продукта A_1 равен c_1, в рацион входит x_1 единиц продукта общей стоимостью $c_1 x_1$. Учитывая затраты на другие виды продуктов, получим стоимость каждого рациона:

$$f = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 \to min\ .$$

Далее определим, какое количество первого вида питательного вещества B_1 содержится в рационе. В табл. 2 приведены данные a_{i1} о содержании вещества B_1 в единице каждого вида продуктов A_i. Эти величины необходимо умножить на x_i и составить сумму: $a_{11} x_1 + a_{21} x_2 + a_{31} x_3 + a_{41} x_4$. Требуется, чтобы в рационе суммарное количество этого вида питательного вещества было не меньше, чем минимальная суточная потребность: $a_{11} x_1 + a_{21} x_2 + a_{31} x_3 + a_{41} x_4 \geq b_1$.

Запишем всю систему ограничений по видам питательных веществ:

$$\begin{cases} a_{11} x_1 + a_{21} x_2 + a_{31} x_3 + a_{41} x_4 \geq b_1\,; \\ a_{12} x_1 + a_{22} x_2 + a_{32} x_3 + a_{42} x_4 \geq b_2\,; \\ a_{13} x_1 + a_{23} x_2 + a_{33} x_3 + a_{43} x_4 \geq b_3\,. \end{cases}$$

К этой системе следует добавить условие неотрицательности неизвестных:

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

Математическая модель задачи представлена в таком виде:

Найти минимум линейной функции цели

$$f = \sum_{i=1}^{m} c_i x_i \to min$$

при ограничениях:

$$\begin{cases} \sum\limits_{i=1}^{m} a_{ij} x_i \geq b_j \\ (j = \overline{1, n}) \end{cases},$$

$$x_i \geq 0.$$

Где n – количество видов питательных веществ, m – количество продуктов, которые входят в рацион. Эта задача есть задачей линейного программирования. По сравнению с предыдущей задачей модели в некотором содержании взаимны – сейчас требуется найти минимум линейной функции цели при линейных ограничениях типа $\geq$. Будем называть такую форму 2-й стандартной формой задачи ЗЛП.

Транспортная задача

В транспортной задаче требуется составить наиболее дешевый план перевозки однородного продукта с пунктов производства A_1, A_2 в пункты потребления B_1, B_2, B_3. В табл. 3.3 представлены сведения о запасах a_i в пунктах производства, спрос b_j в пунктах потребления и тарифы c_{ij} – стоимости перевозки единицы товара с каждого пункта отправления в каждый пункт предназначения:

Таблица 3.3

Условие задачи

Поставщики	Потребители			Запасы
	B_1	B_2	B_3	
A_1	c_{11}	c_{12}	c_{13}	a_1
A_2	c_{21}	c_{22}	c_{23}	a_2
Спрос	b_1	b_2	b_3	

Неизвестные x_{ij} – количество груза, перевезенного от каждого поставщика A_i к каждому потребителю B_j. Как правило, в сбалансированном задании общий спрос $b_1 + b_2 + b_3$ равен общему запасу $a_1 + a_2$. Необходимо вывезти все запасы и удовлетворить весь спрос с наименьшими транспортными затратами.

Составляем функцию цели, описывающую общие транспортные затраты. Тариф c_{11} показывает стоимость перевозки единицы продукции по маршруту A_1–B_1; затраты на перевозку по этому маршруту количества товара x_{11} будут равны произведению $c_{11}x_{11}$. Общие затраты будут равны сумме таких величин:

$$f = (c_{11}x_{11} + c_{12}x_{12} + c_{13}x_{13}) + (c_{21}x_{21} + c_{22}x_{22} + c_{23}x_{23}) \rightarrow min .$$

Количество товара, вывезенного с A_i, должно быть равным запасу a_i:

$$\begin{cases} x_{11} + x_{12} + x_{13} = a_1; \\ x_{21} + x_{22} + x_{23} = a_2. \end{cases}$$

Количество товара, вывезенного с B_j, должно быть равным спросу b_j:

$$\begin{cases} x_{11} + x_{12} = b_1; \\ x_{21} + x_{22} = b_2; \\ x_{31} + x_{32} = b_3. \end{cases}$$

Все неизвестные в этой задаче - неотрицательные:

$$x_{ij} \geq 0.$$

Математическая модель транспортной задачи представлена в таком виде: найти минимум линейной функции цели

$$f = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \rightarrow \min$$

при ограничениях-равенствах

$$\begin{cases} \sum\limits_{j=1}^{n} x_{ij} = a_i \\ (i = \overline{1, m}) \end{cases}; \qquad \begin{cases} \sum\limits_{i=1}^{m} x_{ij} = b_j \\ (j = \overline{1, n}) \end{cases};$$

$$x_{ij} \geq 0.$$

Транспортная задача также является задачей ЛП.

Классическими примерами ЗЛП, также, являются: задача о раскрое материалов, задача о размещении заказа.

3.2. Каноническая форма задачи линейного программирования

В общем задании ЗЛП ищется максимум или минимум линейной функции при наличии системы линейных ограничений в виде равенств и неравенств любого типа ($\le$ или $\ge$). Неизвестные могут быть неотрицательными, не иметь никаких ограничений на знак, иметь двусторонние ограничения типа $d_j \le x_j \le g_j$.

Существует стандартные формы задания, две из них были рассмотрены выше (их называют 1-й и 2-й стандартными формами). Третья (основная) стандартная форма называется «канонической».

В канонической форме ищется максимум линейной функции цели

$$z = \sum_{j=1}^{n} c_j x_j \to max$$

при наличии системы линейных ограничений-равенств:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i \\ (i = \overline{1, m}) \end{cases} .$$

Все неизвестные должны быть неотрицательными: $x_j \ge 0$.

Любую задачу ЛП всегда можно привести к канонической форме.

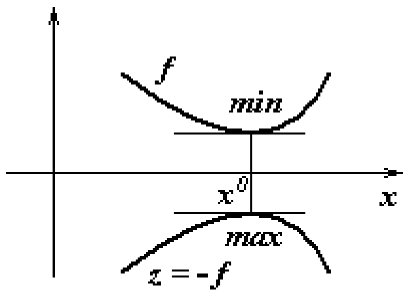


Рис.1. $f_{min}(x^0) = -z_{max}(x^0)$

а) Поиск минимум функции f всегда можно заменить поиском максимума функции $z = -f$, потому что $f_{min}(x^0) = -z_{max}(x^0)$, где x^0 – оптимальное значение неизвестной (см. рис. 1).

б) Ограничения-неравенства типа $\Sigma a_{ij}x_j \le b_i$ можно заменить на равенства $\Sigma a_{ij}x_j + x_{i+k} = b_i$, а неравенства $\Sigma a_{ij}x_j \ge b_i$ – на $\Sigma a_{ij}x_j - x_{i+k} = b_i$. В обоих этих случаях вводятся дополнительные неотрицательные балансовые неизвестные $x_{i+k} \ge 0$, где k – число исходных неизвестных.

в) Как правило, все неизвестные – неотрицательные по своему физическому (экономическому) смыслу. Однако необходимо предусмотреть наличие нескольких неизвестных без любых ограничений на знак – такие неизвестные часто появляются при постановке заданий нелинейного программирования. Неизвестные без ограничения на знак всегда можно заменить разностью двух неотрицательных неизвестных $x_j = x'_j - x''_j$, где $x'_j, x''_j \ge 0$. Этот прием довольно прост, но приводит к увеличению общего числа неизвестных в канонической форме задачи. Другой способ: можно выразить неизвестные без ограничения на знак через другие неотрицательные неизвестные и, таким образом, исключить их с дальнейшего анализа до окончания процесса оптимизацию. Размер задачи при этом не увеличивается, а уменьшается.

При односторонних ограничениях типа $x_j \geq d_j$ вводим новые неотрицательные неизвестные $x'_j = x_j - d_j$. Ограничения типа $x_j \leq g_j$ включаем в общее число линейных ограничений, которые преобразовываются к виду равенств путем введения дополнительных балансовых неизвестных.

Кроме неизвестных без ограничения на знак иногда приходиться вводить фиктивные неизвестные, тождественно равные нолю $x_j \equiv 0$. Далее будет доказано, что ограничения-равенства, фиктивные неизвестные и неизвестные без ограничения на знак тесно связаны между собой.

В канонической форме m – общее число ограничений-равенств, , n – общее число (неотрицательных) неизвестных, включая балансовые.

3.3. Графическое решение задач математического программирования

Геометрическая интерпретация целесообразна, если многоугольник планов принадлежит двумерному пространству. В случае трехмерного пространства задача усложняется, а для четырехмерного графическое решение становится невозможным. Пусть задача имеет вид:

$$Z = c_1 x_1 + c_2 x_2 \quad \rightarrow \quad \max \tag{3.1}$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \leq b_1 \\ a_{21}x_1 + a_{22}x_2 \leq b_2 \\ \dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 \leq b_m \end{cases} \tag{3.2}$$

$$x_1 \geq 0, \quad x_2 \geq 0. \tag{3.3}$$

Каждое неравенство согласно с уравнением прямой $a_{i1}x_1 + a_{i2}x_2 = b_i$ разделяет плоскость $0X_1X_2$ на две полуплоскости и определяет ту полуплоскость, в которой выполняется неравенство (Рис.2). Ограничения по знаку (3.3) также определяют полуплоскости, ограниченные прямыми: x_1= 0 та x_2= 0; таким образом, область определения данной задачи размещена в I-й четверти координатной плоскости. Если система ограничений совместны, полуплоскости, где выполняются соответствующие неравенства, пересекаются образуют многоугольник возможных решений ЗЛП, т.е. многоугольник планов (рис. 3).

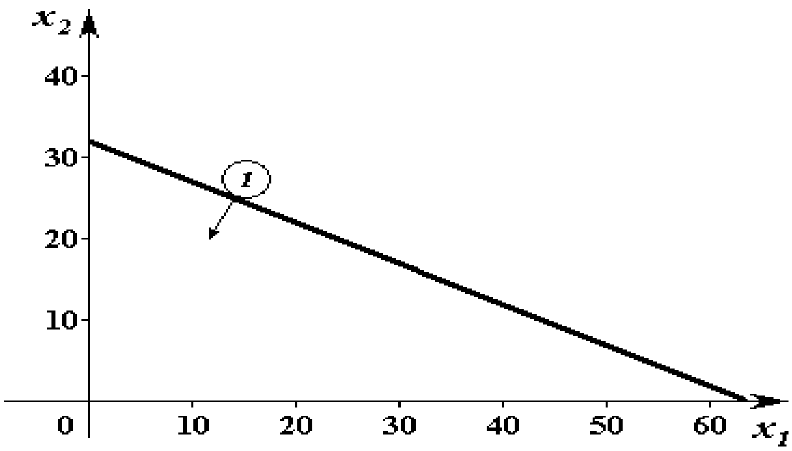


Рис. 2. Область решения 1–го неравенства

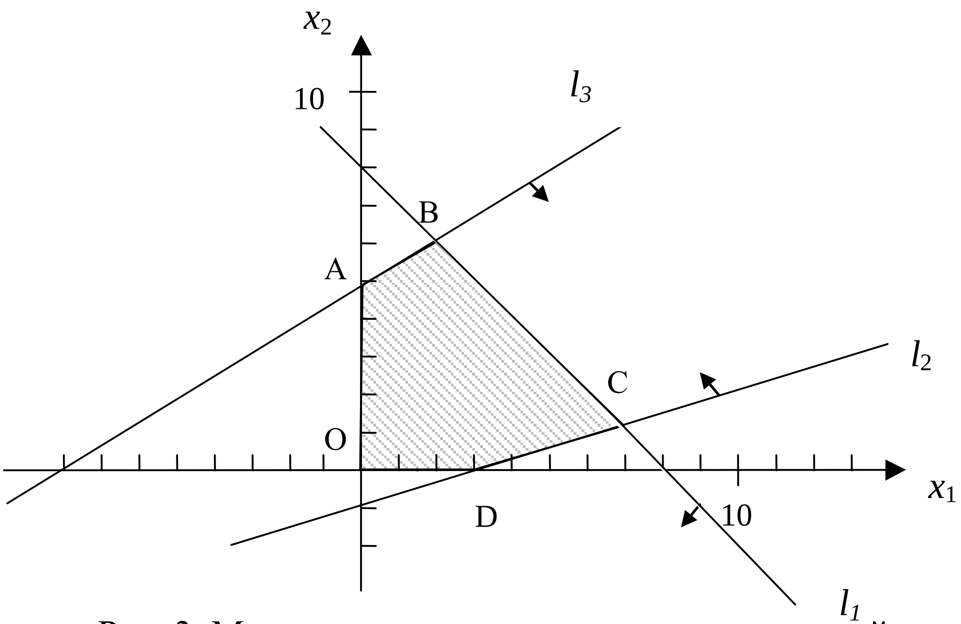


Рис. 3. Многоугольник планов, соответствующий первоначальной системе

Рассмотрим графическую интерпретацию целевой функции. Поскольку целевая функция есть функцией двух переменных, на плоскости $0X_1X_2$ изображается линией уровня, т.е. линией, соответствующей постоянному значению целевой функции Z = Const. Поскольку, линия уровня, соответствующая нулевому значению целевой функции, имеет уравнение: $c_1x_1 + c_2x_2 = 0$, ее графиком есть прямая проходящая через начало координат. Коэффициенты целевой функции имеют значения проекций нормали к линии уровня $\mathbf{N} = (c_1; c_2)$. Для линейной функции градиент, т.е. вектор, указывающий направление наискорейшего роста функции, совпадает с вектором нормали: $\mathbf{gradZ} = \frac{\partial z}{\partial x_1}\mathbf{i} + \frac{\partial z}{\partial x_2}\mathbf{j} = c_1\mathbf{i} + c_2\mathbf{j} = \mathbf{N}$. Если значение целевой функции возрастает, линия уровня передвигается в направлении нормали (рис. 3).

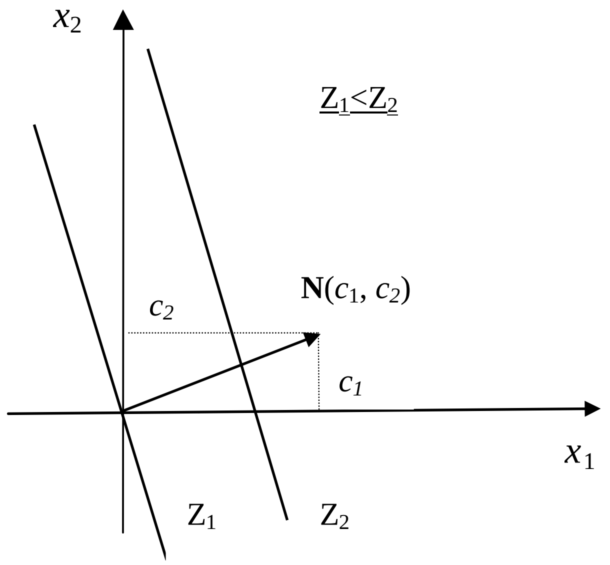


Рис. 4. Построение линии уровня

Замечание. Аналогичную графическую интерпретацию может иметь задача линейного программирования и для большего количества переменных, если разность между количеством переменных и количеством уравнений основной системы ограничений равна двум: $n - m = 2$.

Анализируя графическое решение, имеем общие выводы: 1) система линейных ограничений определяет выпуклую многоугольную область планов; 2) оптимальное решение ЗЛП достигается в одной из вершин многоугольной области планов. Имеют место следующие случаи:

1. Система ограничений противоречивая, область планов – пустая ($D = \varnothing$); в этом случае задача решения не имеет (рис. 5).

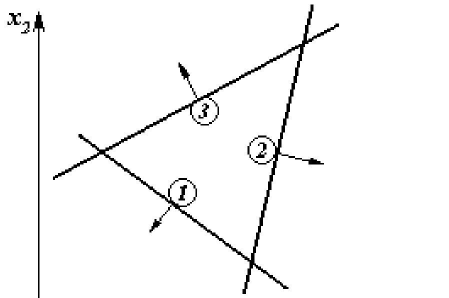


Рис. 5. Нет решений

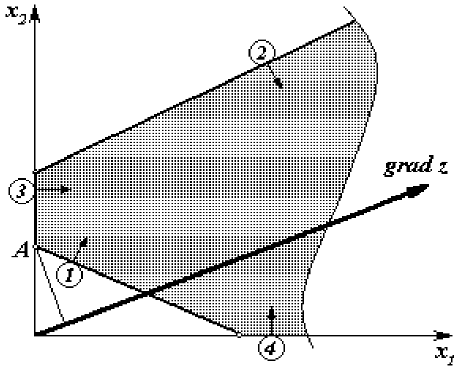
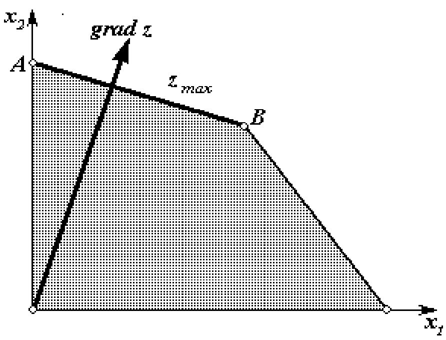


Рис.6. Неограниченная область планов



2. Область планов открытая, а максимальное значение функции цели – неограниченное (рис. 6).

3. Бесконечное множество эквивалентных оптимальных планов (рис. 7)

4. Область планов может быть открытой, а предельные значения функции цели – неограниченными (рис. 8).

Рис.7. Неединственность оптимального решения

5. Область планов может выродиться в единственную точку. Это единственное решение в силу отсутствия выбора будет оптимальным (рис. 9).

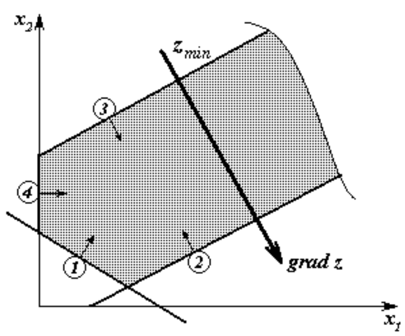


Рис.8. Область планов открытая, функция цени - ограничена

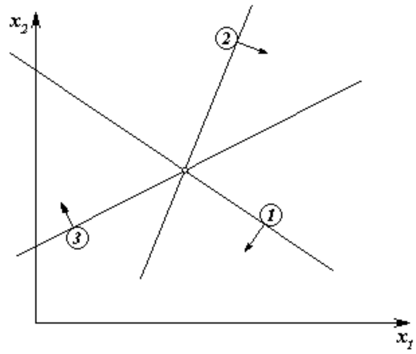


Рис.9. Единственное решение

6. Область планов может представляться линией (отрезок) (рис. 10).

7. Особый случай вырождения (в вершине, где оптимальное значение – пересекаются три линии) (рис. 11).

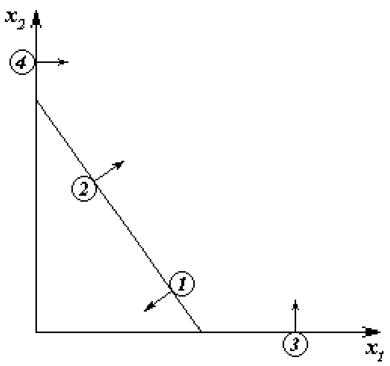


Рис.10. Область решений – отрезок

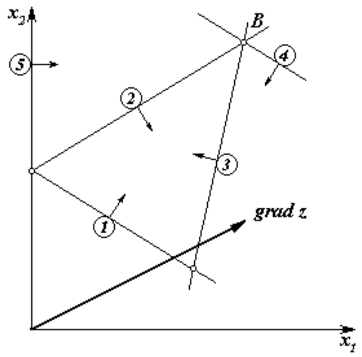


Рис.11. Вырождение плана

3.4. Свойства возможных решений задачи линейного программирования

Рассмотрим основную систему ограничений ЗЛП в канонической форме:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad i = \overline{1,m}. \tag{3.4}$$

Постановка ЗЛП имеет смысл только когда система (3.4) имеет множество решений. В этом случае среди множества решений мы выбираем то, при котором целевая функция достигает экстремума. Согласно теореме Кронекера – Капелли, система m уравнений с n неизвестными совместна и имеет множество решений, если ранг матрицы коэффициентов системы $\mathbf{A} = (a_{ij})_{m\,x\,n}$ равен рангу расширенной матрицы A|B и меньше количества неизвестных:

$$\begin{gathered} r_A = r_{A|B} = k \\ r_A = r_{A|B} < n \end{gathered}.$$

Неизвестные, коэффициенты при которых образуют матрицу порядка k, и определитель которой не равен нолю, является основными (или базисными). Остальные $n - k$ неизвестных есть свободными. Если в общем решении системы, где основные неизвестные определяются через свободные, свободным неизвестным присвоить значение равное нулю, то такое решение называется базисным. Среди множества решений основной системы ограничений (2.4) количество базисных решений не превышает количества способов, которыми можно отобрать m переменных среди n их общего количества, т.е. C_n^m - количество сочетаний из n по m. Если принять во внимание ограничения на знак:

$$x_j \ge 0, \quad j = \overline{1,m}, \tag{3.5}$$

то среди базисных решений планами ЗЛП будут только те, для которых все базисные переменные неотрицательные. Такие планы называются опорными. Если все k базисных компонентов являются положительными, соответствующий план называется невырожденным. В противном случае (то есть, когда количество положительных компонентов опорного плана меньше чем k) план есть вырожденным.

***Теорема* 3.1.** Множество планов задачи линейного программирования является выпуклым.

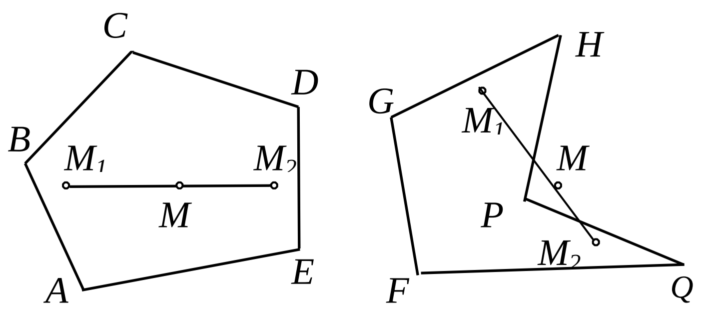


Рис. 12. Выпуклое (*ABCDE*) и невыпуклое (*FGYPQ*) множества

Заметим, что выпуклым есть множество, для которого выполняется такое утверждение: если точки M_1 и M_2 принадлежат этому множеству, то любая внутренняя точка M отрезка M_1M_2 также принадлежит этому множеству. То есть $M = \lambda M_1 + (1-\lambda)M_2$, $0 \le \lambda \le 1$, где M_1, M_2, M − радиус-вектора, что соответствуют этим точкам. Говорят, что вектор M является выпуклой линейной комбинацией векторов M_1 и M_2.

Доказательство. Рассмотрим задачу линейного программирования в канонической форме:

$$Z = \boldsymbol{C} \cdot \boldsymbol{X} \quad \to \quad \max$$

$$\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{B} \tag{3.6}$$

$$\boldsymbol{X} \ge 0. \tag{3.7}$$

Пусть вектора $\boldsymbol{X}_1$ и $\boldsymbol{X}_2$ – возможные решения задачи, то есть удовлетворяют основной системе ограничений (3.6):

$$\boldsymbol{A} \cdot \boldsymbol{X}_1 = \boldsymbol{B} \quad и \quad \boldsymbol{A} \cdot \boldsymbol{X}_2 = \boldsymbol{B}.$$

Рассмотрим вектор X, который является линейной комбинацией векторов X_1 и X_2:

$$\boldsymbol{X} = \lambda \boldsymbol{X}_1 + (1-\lambda)\boldsymbol{X}_2, \quad 0 \le \lambda \le 1,$$

и подставим его в систему ограничений (3.6):

$$\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{A} \cdot (\lambda \boldsymbol{X}_1 + (1-\lambda)\boldsymbol{X}_2) = \lambda \cdot \boldsymbol{A}\boldsymbol{X}_1 + \boldsymbol{A}\boldsymbol{X}_2 - \lambda \cdot \boldsymbol{A}\boldsymbol{X}_2 =$$
$$= \lambda \boldsymbol{B} + \boldsymbol{B} - \lambda \boldsymbol{B} = \boldsymbol{B}.$$

То есть, вектор X также есть решением основной системы ограничений и принадлежит к множеству возможных решений задачи линейного программирования – что и требовалось доказать.

Если рассмотреть графическую интерпретацию выпуклого множества (рис.12), то выпуклым есть пятиугольник *ABCDE*, наоборот, пятиугольник

FGHPQ не есть выпуклым, поскольку не все точки отрезка M_1M_2 принадлежат этому множеству.

Согласно теореме 3.1 только пятиугольник *ABCDE* может быть многоугольником планов задачи линейного программирования.

***Теорема* 3.2.** Опорное решение задачи линейного программирования соответствует допустимой экстремальной точке пространства решений (без доказательства).

Другими словами, каждому опорному плану задачи линейного программирования соответствует вершина (угловая точка) многоугольника планов и, наоборот, каждой вершине многоугольника планов соответствует опорный план.

Вопросы для самопроверки:

1) Какие типовые задачи линейного программирования различают?
2) Записать задачу линейного программирования в канонической форме.
3) Какие задачи линейного программирования можно решить графически?
4) Каковы основные этапы графического метода решения задач линейного программирования?
5) Какие случаи графического решения задачи линейного программирования различают?
6) Привести свойства возможных решений задачи линейного программирования.

Тема 4. Симплексный метод решения задач линейного программирования и некоторые его теоретические аспекты

4.1. Поиск оптимального плана. Условие оптимальности

4.2. Алгоритм симплексного метода

4.3. Метод искусственного базиса. Расширенная М-задача

4.4. Проблема вырождения

4.1. Поиск оптимального плана. Условие оптимальности

Рассмотрим задачу линейного программирования в канонической форме:

$$Z = \sum_{i=1}^{n} c_j x_j \quad \to \quad \max$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{1,m}$$

$$x_j \ge 0, \quad j = \overline{1,n}$$

Чтобы задача имела решение, система ее ограничений должна быть совместной. Это возможно, если ранг r системы не больше числа неизвестных n: $r \le n$. Случай $r > n$ невозможен. При $r = n$ системы имеет единственное решение. Если же целевая функция достигает экстремального значения более чем в одной крайней точке, то она достигает того же значения в любой точке, которая является их выпуклой линейной комбинации. Для решения задачи линейного программирования (ЗЛП) необходимо найти все крайние точки многогранника планов (из не больше чем C_n^r) и сравнить в них значение целевой функции. Общая идея симплекс-метода (метода последовательного улучшения плана) состоит в 1) нахождении начального опорного плана; 2) наличия признака оптимальности опорного плана; 3) умения переходить к не худшему опорному плану.

Рассмотрим теоретические основы симплекс-метода. С помощью метода Жордана-Гаусса выразим первые неизвестные в виде:

$$\begin{cases} x_1 + a_{1\,m+1}x_{m+1} + \ldots + a_{1\,n}x_n = b_1 \\ x_2 + a_{2\,m+1}x_{m+1} + \ldots + a_{2\,n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m + a_{m\,m+1}x_{m+1} + \ldots + a_{m\,n}x_n = b_m \end{cases},$$

Т.е. имеем m базисных неизвестных: $x_1, x_2, \ldots, x_m$ выразим их через свободные неизвестные $x_{m+1}, x_{m+2}, \ldots, x_n$:

$$x_i = b_i - \sum_{k=1}^{n-m} a_{i\,k}x_{m+k}, \quad i = \overline{1,m}.$$

Получим исходный опорный план:

$$\mathbf{X_0} = (b_1, b_2, \ldots, b_m, \underbrace{0, 0, \ldots, 0}_{n-m}).$$

Значение функции для исходного опорного плана:

$$Z_o = Z(\mathbf{X_0}) = c_1x_1 + c_2x_2 + \ldots + c_mx_m.$$

Базис составляют m векторов.

$$\mathbf{A_1} = \begin{pmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix}, \quad \mathbf{A_2} = \begin{pmatrix} 0 \\ 1 \\ \ldots \\ 0 \end{pmatrix}, \ldots, \mathbf{A_m} = \begin{pmatrix} 0 \\ 0 \\ \ldots \\ 1 \end{pmatrix}.$$

$$\mathbf{A}_{m+1} = \begin{pmatrix} a_{1\,m+1} \\ a_{2\,m+1} \\ \ldots\ldots \\ a_{m\,m+1} \end{pmatrix}, \quad \mathbf{A}_{m+2} = \begin{pmatrix} a_{1\,m+2} \\ a_{2\,m+2} \\ \ldots\ldots \\ a_{m\,m+2} \end{pmatrix}, \quad \ldots, \quad \mathbf{A}_n = \begin{pmatrix} a_{1\,n} \\ a_{2\,n} \\ \ldots\ldots \\ a_{m\,n} \end{pmatrix}, \quad \mathbf{A}_o = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}$$

$$\begin{cases} \mathbf{A}_{m+1} = a_{1m+1}\mathbf{A}_1 + a_{2m+1}\mathbf{A}_2 + \ldots + a_{mm+1}\mathbf{A}_m \\ \mathbf{A}_{m+2} = a_{1m+2}\mathbf{A}_1 + a_{2m+2}\mathbf{A}_2 + \ldots + a_{mm+2}\mathbf{A}_m \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \mathbf{A}_n = a_{1n}\mathbf{A_1} + a_{2n}\mathbf{A_2} + \ldots + a_{mn}\mathbf{A_m} \\ \mathbf{A_o} = b_1\mathbf{A_1} + b_2\mathbf{A_2} + \ldots + b_m\mathbf{A_m} \end{cases} \qquad (4.1)$$

Для уменьшения значения целевой функции надо построить другой опорный план, т.е. перейти к новому базису.

Введем в базис вектор $\mathbf{A}_{m+1}$. Для этого в выражении (4.1) умножим вектор $\mathbf{A}_{m+1}$ на некоторое положительное число $\theta \geq 0$ и вычтем его с вектора $\mathbf{A}_o$. Имеем

$$\mathbf{A}_{m+1}\theta = a_{1m+1}\theta\mathbf{A}_1 + a_{2m+1}\theta\mathbf{A}_2 + \ldots + a_{mm+1}\theta\mathbf{A}_m$$

$$A_0 - A_{m+1}\theta = b_1\mathbf{A_1} + b_2\mathbf{A_2} + \ldots + b_m\mathbf{A_m} - (a_{1m+1}\theta\mathbf{A}_1 + a_{2m+1}\theta\mathbf{A}_2 + \ldots + a_{mm+1}\theta\mathbf{A}_m)$$

$$\mathbf{A_o} - \theta \cdot \mathbf{A}_{m+1} =$$
$$= (b_1 - \theta \cdot a_{1\,m+1})\mathbf{A}_1 + (b_2 - \theta \cdot a_{2\,m+1})\mathbf{A}_2 + \ldots + (b_m - \theta \cdot a_{m\,m+1})\mathbf{A}_m$$

или:

$$\mathbf{A}_o =$$
$$= \theta \cdot \mathbf{A}_{m+1} + (b_1 - \theta \cdot a_{1m+1})\mathbf{A}_1 + (b_2 - \theta \cdot a_{2m+1})\mathbf{A}_2 + \ldots + (b_m - \theta \cdot a_{mm+1})\mathbf{A}_m \qquad (4.2)$$

Выражение (4.2) можно рассматривать как разложение вектора $\mathbf{A}_o$ за новым базисом. Один с коэффициентов разложения должен раняться нулю.

$$b_i - \theta \cdot a_{i\,m+1} = 0.$$

Выберем его по критерию:

$$\theta = \min\left\{ \frac{b_i}{a_{i\,m+1}} \geq 0, \quad i = \overline{1,m} \right\}.$$

Число θ будем называть симлексным отношением. Для исходного плана $\mathbf{X}_o$ вектор $\mathbf{C}_б$ равен:

$$\mathbf{C}_б = (c_1, c_2, \ldots, c_m).$$

Соответственно исходное значение целевой функции можно определить так:

$$Z_o = Z(\mathbf{X}_o) = c_1 b_1 + c_2 b_2 + \ldots + c_m b_m = \mathbf{C}_б \cdot \mathbf{A}_o$$

Обозначим через Z_j скалярное произведение вектора $\mathbf{C}_б$ на произвольный вектор $\mathbf{A}_j$:

$$Z_j = \mathbf{C}_б \cdot \mathbf{A}_j, \quad j = \overline{1,m}$$

И рассмотрим разность:

$$\Delta_j = Z_j - c_j, \quad j = \overline{1,m},$$

которая называется оценкой плана за j-м вектором относительно базиса.

Следовательно, для всех базисных векторов оценки плана должны равняться нулю. Аналогично вычисляют оценки для векторов, не вошедших в базис:

$$\Delta_{m+1} = Z_{m+1} - c_{m+1} = \mathbf{C}_б \cdot \mathbf{A}_{m+1} - c_{m+1} = (c_1, c_2, ..., c_m) \cdot \begin{pmatrix} a_{1m+1} \\ a_{2m+1} \\ ... \\ a_{mm+1} \end{pmatrix} - c_{m+1} =$$

$$= c_1\, a_{1m+1} + c_2\, a_{2m+1} + ... + c_m\, a_{mm+1} - c_{m+1}$$

Для этих векторов оценки плана не равны нулю.

Теорема. Если для некоторого опорного плана $\mathbf{X}_o$ задачи линейного программирования при нахождении максимума целевой функции существует хотя бы одна отрицательная оценка $\Delta_j < 0$, то этот план не является оптимальным, и в этом случае можно найти новый план $\mathbf{X}_1$, при котором целевая функция будет иметь значение не меньше исходного:

$$Z(\mathbf{X}_1) \geq Z(\mathbf{X}_o).$$

Следствие. Если для некоторого опорного плана задачи линейного программирования на исследование максимального значения целевой функции оценки всех векторов $\Delta_j \geq 0 \quad (j = \overline{1,m})$ относительно базиса неотрицательные, то данный опорный план является оптимальным.

4.2. Алгоритм симплексного метода

Алгоритм симплексного метода состоит из таких этапов:

1. Вычисляют при исходном опорном плане все значения строки Z. Если среди них есть положительные, то план не является оптимальным и его следует улучшить.

2. Чтобы улучшить план, в базис вводят один из векторов, который имеет положительное значение Z. Среди них выбирают тот, который имеет наибольшее значение Z.

3. Выводят из базиса той вектор, который имеет минимальное неотрицательное симплексное отношение:

$$\theta = \min \left\{ \frac{b_i}{a_{i\,m+1}} \geq 0, \quad i = \overline{1,m} \right\}.$$

4. Переход к новому опорному плану $\mathbf{X}_1$ осуществляется по методу Жордана - Гаусса. Итерации выполняются до тех пор, пока значения Z станут все неположительными, что отвечает оптимальному плану.

5. С основной системы ограничений, которая отвечает последней итерации, находят компоненты оптимального плана $\mathbf{X}^* = (x_1^*, x_2^*, \ldots, x_m^*)$.

6. По оптимальному плану $\mathbf{X}^*$ вычисляют экстремальное значение целевой функции $Z = Z(\mathbf{X}^*)$.

Замечание. Если некоторому вектору $\mathbf{A}_j$ соответствует отрицательная оценка относительно рассматриваемого опорного плана, а сам вектор не имеет положительных компонент в данном базисе, то задача линейного программирования при определении максимального значения целевой функции не имеет решений.

4.3. Метод искусственного базиса. Расширенная М-задача

Пусть имеем

$$Z = \sum_{j=1}^{n} c_j x_j \quad \rightarrow \quad extr: \max \textit{ или } \min$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le b_k \\ a_{k+11}x_1 + a_{k+12}x_2 + \ldots + a_{k+1n}x_n \ge b_{k+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{l1}x_1 + a_{l2}x_2 + \ldots + a_{ln}x_n \ge b_l \\ a_{l+11}x_1 + a_{l+12}x_2 + \ldots + a_{l+1n}x_n = b_{l+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \\ \quad x_j \ge 0, \quad j = \overline{1,n} \end{cases}.$$

добавим в первые k неравенства $x_{n+1}, \ldots, x_{n+k}$,

$\boldsymbol{A}_{n+1}, \ldots, \boldsymbol{A}_{n+k}$ составят базис.

$$Z = c_1x_1 + c_2x_2 + c_nx_n + 0 \cdot x_{n+1} + \ldots + 0 \cdot x_{n+l} \quad \rightarrow \quad extr$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} = b_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + x_{n+k} = b_k \\ a_{k+11}x_1 + a_{k+12}x_2 + \ldots + a_{k+1n}x_n - x_{n+k+1} = b_{k+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{l1}x_1 + a_{l2}x_2 + \ldots + a_{ln}x_n - x_{n+l} = b_l \\ a_{l+11}x_1 + a_{l+12}x_2 + \ldots + a_{l+1n}x_n = b_{l+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \\ \quad x_j \ge 0, \quad j = \overline{1,n+l} \end{cases}.$$

Векторы $\boldsymbol{A}_{n+1}, \ldots, \boldsymbol{A}_{n+l}$ не образуют базис.

Используем метод искусственного базиса. Идея его состоит с следующем: в левую часть уравнения добавляют фиктивную переменную. Если целевая функция исследуется на минимуму в нее включает искусственная переменная с очень большим положительным коэффициентом M, если на максимум – M отрицательное, т.е. очень мало. Базис, содержащий искусственную переменную, называется искусственным, а задача линейного программирования называется расширенной или M-задачей.

Пример. Известно, что по агротехническим нормам в 1 га чернозема надо вносить в течении сезона основные полезные вещества в количествах не менее 6 ед. фосфора, 8 ед. азота, 12 ед. калия. Фермерское хозяйство имеет возможность приобрести удобрений состава A_1, A_2 и A_3. В 1 кг удобрений в зависимости от состава содержится основных полезных веществ в количествах:

A_1:	2 ед. P,	1 ед. N,	3 ед. K
A_2:	1 ед. P,	2 ед. N,	4 ед. K
A_3:	3 ед.. P,	1,5 ед. N,	2 ед. K

Цены за 1кг удобрений: A_1 – 2 грн., A_2 – 3,5 грн., A_3 – 2,5 грн. Составить наиболее экономный план закупки удобрений.

Решение:

$$Z = 2x_1 + 3{,}5x_2 + 2{,}5x_3 \quad \to \min .$$

$$\begin{cases} 2x_1 + x_2 + 3x_3 \ge 6 \\ x_1 + 2x_2 + 1{,}5x_3 \ge 8 \\ 3x_1 + 4x_2 + 2x_3 \ge 12 \\ \quad x_j \ge 0, \quad j = \overline{1,3} \end{cases}.$$

$$Z = 2x_1 + 3{,}5x_2 + 2{,}5x_3 + 0 \cdot y_1 + 0 \cdot y_2 + 0 \cdot y_3 \quad \to \min$$

$$\begin{cases} 2x_1 + x_2 + 3x_3 - y_1 = 6 \\ x_1 + 2x_2 + 1{,}5x_3 - y_2 = 8 \\ 3x_1 + 4x_2 + 2x_3 - y_3 = 12 \\ \quad x_j \ge 0, \quad j = \overline{1,6} \end{cases}.$$

$$\begin{cases} x_1 + 3x_2 - x_3 + y_1 - y_3 = 6 \\ 2x_1 + 2x_2 + 0{,}5x_3 + y_2 - y_3 = 4 \\ 3x_1 + 4x_2 + 2x_3 - y_3 = 12 \\ \quad x_j \ge 0, \quad j = \overline{1,6} \end{cases}.$$

$$Z = 2x_1 + 3{,}5x_2 + 2{,}5x_3 + 0 \cdot y_1 + 0 \cdot y_2 + 0 \cdot y_3 + M \cdot y_4 \quad \to \min$$

$$\begin{cases} x_1 + 3x_2 - x_3 + y_1 - y_3 = 6 \\ 2x_1 + 2x_2 + 0{,}5x_3 + y_2 - y_3 = 4 \\ 3x_1 + 4x_2 + 2x_3 - y_3 + y_4 = 12 \\ \quad x_j \ge 0, \quad j = \overline{1,7} \end{cases}.$$

№	Базис	A_0	c1=2	c2=3,5	c3=2,5	c4=0	c5=0	c6=0	c7=M
			X 1	X 2	X 3	Y1	Y2	Y3	Y4
1	Y1	6	1	3	-1	1	0	-1	0
2	**Y2**	4	2	2	1/2	0	1	-1	0
3	Y4	12	3	4	2	0	0	-1	1
Z		12M	3**M**	4M	2M	0	0	-M	M
$\Delta_j = Z_j - c_j$			3M-2	4M- -3,5	2M- -2,5	0	0	-M	0

$\theta_2 = \min\left\{6/3;\ 4/2;\ 12/4\right\} = 2.$

№	Базис	A_0	X 1	X 2	X 3	Y1	Y2	Y3	Y4
4	Y1	0	-2	0	-7/4	1	-3/2	1/2	0
5	**X2**	2	1	1	1/4	0	1/2	-1/2	0
6	Y4	4	-1	0	1	0	-2	1	1
Z		4M+ +7	7/2- -M	7/2	7/8+ +M	0	7/4- -2M	M-7/4	M
$\Delta_j = Z_j - c_j$			3/2- -M	0	M- -13/8	0	7/4- -2M	**M-7/4**	0

7	Y1	7	-15/4	0	0	1	-5	9/4	7/4
8	**X2**	1	5/4	1	0	0	1	-3/4	-1/4
9	X3	4	-1	0	1	0	-2	1	1
Z		13,5	15/8	7/2	5/2		-3/2	-1/8	13/8
$\Delta_j = Z_j - c_j$			-1/8	0	0	0	-3/2	-1/8	13/8- -M

$X^* = (0;\ 1;\ 4;\ 7;\ 0;\ 0;\ 0)$.

Таким образом, для 1 га чернозема, фермерскому хозяйству следует закупить 1 кг удобрений состава A_2 и 4 кг удобрений состава A_3 (удобрение состава A_1 совсем не закупать). Этим полностью удовлетворятся требования внесения в чернозем азота и калия, а количество фосфора на 1 га превысит необходимый минимум на 7 единиц. Общая стоимость удобрений для этого плана минимальная и составит 13,5 грн с расчета 1 га.

4.4. Проблема вырождения

При использовании симплексного метода монотонное возрастание целевой функции при исследовании на максимум (или монотонное уменьшение при исследовании на минимум) имеет место при условии, что на каждой итерации мы получаем невырожденный опорный план. Имеют место задачи, когда основная система ограничений в правой части содержит один или несколько нулей. Рассмотрим, как изменится целевая функция, если при очередном преобразовании по методу Жордана - Гаусса элемент решения содержится в строке с нулевой правой частью. Известно, что при каждой итерации целевая функция изменяется на величину $\Delta Z = |\theta \cdot \Delta|$. Поскольку в нашем случае $\theta = 0$, то введение в базис вектора, которому соответствует нулевое симплексное отношение, не приводит к изменению целевой функции, следовательно, такое преобразование не имеет смысла. Как правило, после нескольких преобразований с использованием нулевого симплексного отношения можно получить план, который был ранее, то есть имеет место зацикливание.

На практике такие случаи встречаются очень редко, и имеют, как правило, искусственный характер, но теоретически такой случай возможен.

Вопросы для самопроверки:

1) Для решения каких математических задач используется симплексный метод?
2) В чем суть симплекс-метода?
3) На каких свойствах задач линейного программирования основан симплекс-метод?
4) Как определить первоначальный опорный план задачи линейного программирования?
5) Сформулируйте последовательность этапов практической реализации симплекс-метода при решении задач линейного программирования.
6) Какие признаки оптимальности базисного плана?
7) Когда возникает необходимость использования симплекс-метода с искусственным базисом?
8) В чем суть проблемы вырождения?

Тема 5. Теория двойственности. Взаимно двойственные задачи линейного программирования

5.1. Правила составления условий взаимно двойственных задач

5.1. Правила составления условий взаимно двойственных задач

В различных разделах математики часто встречаются так называемые *теоремы двойственности*. Каждая из них позволяет для любого утверждения данной теории построить – по определенным стандартным правилам – другое утверждение таким образом, что из справедливости первого автоматически следует справедливость второго. Применительно к линейному программированию это выражается в том, что, решая одну оптимизационную задачу, мы получаем решение еще одной оптимизационной задачи, *двойственной к исходной*.

Можно указать на ряд практических приложений теории двойственности.

- При решении задачи ЛП симплекс–методом решение парной задачи получается автоматически без дополнительных затрат с нашей стороны, поэтому из двух связанных между собой задач решать следует ту, которая проще.
- Рассматривая по шагам процесс решения исходной задачи стандартным симплекс–методом и наблюдая за связанным с ним процессом автоматического решения двойственной задачи, был открыт алгоритм так называемого *двойственного симплекс–метода*. Новый метод существенно расширил вычислительные возможности и оказался очень полезным при решении целочисленных задач ЛП и задач ЛП с параметрами, о чем будет сказано ниже в соответствующих разделах.
- Интересно, что, записав условия парных задач вместе, оказалось возможным свести задачу оптимизации к обычному решению системы алгебраических уравнений (с двойным числом неизвестных, из которых половина неизвестных должна равняться нулю). Такой прием используется при решении транспортной задачи ЛП и является основой для решения задач квадратичного программирования.

- Решение двойственной задачи полезно само по себе – оно дает возможность провести анализ устойчивости решения исходной задачи по отношению к малым изменениям в ее условиях. Так, в задаче оптимального использования ограниченных ресурсов возможно проверить чувствительность найденного оптимального плана к малым вариациям в запасах ресурсов. Экономическая интерпретация двойственных оценок позволяет определить полезность каждого ресурса и указать наиболее предпочтительный путь изменения их запасов для достижения максимальной прибыли.

Первая задача: предприятие имеет m видов ресурсов в количествах b_i $(i=\overline{1,m})$, с которых изготовляется n видов продукции. Известна матрица технологических коэффициентов $\mathbf{A}=(a_{ij})_{m\,x\,n}$, а также известна стоимость единицы каждой продукции c_j $(j=\overline{1,n})$. Необходимо составить оптимальный план производства, чтобы общая прибыль была наибольшей.

$$Z=c_1x_1+c_2x_2+...+c_nx_n\to\max,$$

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

где $$x_j\ge 0,\quad j=\overline{1,n}.$$

Необходимо оценить имеющиеся ресурсы предприятия, в зависимости от прибыли, полученной от их использования.

Вторая задача: какова должна быть цена единицы каждого из ресурсов, чтобы при заданных количествах ресурсов b_i и величинах стоимости единицы продукции c_j минимизировать общую стоимость затрат. Переменные y_i называются оценками или учетными, неявными ценами.

$$F=b_1y_1+b_2y_2+...+b_my_m\quad\to\quad\min,$$

$$\begin{cases} a_{11}y_1+a_{21}y_2+...+a_{m1}y_m\ge c_1; \\ a_{12}y_1+a_{22}y_2+...+a_{m2}y_m\ge c_2; \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{1n}y_1+a_{2n}y_2+...+a_{mn}y_m\ge c_n; \end{cases}$$

$$y_i\ge 0,\quad i=\overline{1,m}.$$

Обе задачи образуют пару сопряженных двойственных задач.

Правила составления условий взаимно двойственных задач:

1) основным неизвестным одной задачи соответствуют балансовые неизвестные другой задачи и наоборот:

Задача 1	Основные переменные $x_1, x_2, \ldots, x_n$	Балансовые переменные $u_1, u_2, \ldots, u_m$
Задача 2	$v_1, v_2, \ldots, v_n$ Балансовые переменные	$y_1, y_2, \ldots, y_m$ Основные переменные

Число основных неизвестных одной задачи равняется числу ограничений другой задачи.

2) Неотрицательным неизвестным одной задачи соответствуют неотрицательные неизвестные другой задачи.

3) Фиктивным неизвестным (тождественно равным нулю) одной задачи соответствуют в другой задаче неизвестные без ограничения на знак и наоборот.

4) Если условия одной задачи приведены к 1-й стандартной форме, то условия другой задачи будут представлены во 2-й стандартной форме и наоборот.

5) Свободные члены в ограничениях одной задачи являются коэффициентами целевой функции другой задачи и наоборот.

6) Матрицы коэффициентов при неизвестных в ограничениях обеих задач – взаимно транспонированные.

Виды математических моделей двойственных задач:

I. Симметрические задачи:

Исходная задача	Двойственная задача
I.1. $Z = \boldsymbol{C} \cdot \boldsymbol{X} \to \min$	$F = \boldsymbol{Y} \cdot \boldsymbol{A}_0^{\boldsymbol{T}} \to \max$
$\boldsymbol{A} \cdot \boldsymbol{X} \ge \boldsymbol{A}_0$	$\boldsymbol{Y} \cdot \boldsymbol{A} \le \boldsymbol{C}^{\boldsymbol{T}}$
$\boldsymbol{X} \ge \boldsymbol{0}$	$\boldsymbol{Y} \ge \boldsymbol{0}$

Исходная задача	Двойственная задача
I.2. $Z = \boldsymbol{C} \cdot \boldsymbol{X} \to \max$	$F = \boldsymbol{Y} \cdot \boldsymbol{A}_0^{\boldsymbol{T}} \to \min$
$\boldsymbol{A} \cdot \boldsymbol{X} \le \boldsymbol{A}_0$	$\boldsymbol{Y} \cdot \boldsymbol{A} \ge \boldsymbol{C}^{\boldsymbol{T}}$
$\boldsymbol{X} \ge \boldsymbol{0}$	$\boldsymbol{Y} \ge \boldsymbol{0}$

II. Несимметрические задачи:

Исходная задача	Двойственная задача
II.1. $Z = \boldsymbol{C} \cdot \boldsymbol{X} \to \min$	$F = \boldsymbol{Y} \cdot \boldsymbol{A}_0^{\boldsymbol{T}} \to \max$
$\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{A}_0$	$\boldsymbol{Y} \cdot \boldsymbol{A} \le \boldsymbol{C}^{\boldsymbol{T}}$
$\boldsymbol{X} \ge \boldsymbol{0}$	

Исходная задача	Двойственная задача
II.2. $Z = \boldsymbol{C} \cdot \boldsymbol{X} \to \max$	$F = \boldsymbol{Y} \cdot \boldsymbol{A}_0^{\boldsymbol{T}} \to \min$

$$\boldsymbol{A}\cdot\boldsymbol{X}=\boldsymbol{A}_0 \qquad\qquad \boldsymbol{Y}\cdot\boldsymbol{A}\ge\boldsymbol{C}^T$$
$$\boldsymbol{X}\ge\boldsymbol{0}$$

5.2. Теоремы двойственности

Свойства решений взаимно двойственных обычно группируются в два блока, которые именуются 1-й и 2-й теоремами двойственности.

В 1-ю теорему вошли следующие утверждения:

Теорема 1.1. Если одна задача имеет оптимальное решение, то другая задача также имеет оптимальное решение, причем $Z_{\max}(\boldsymbol{X}^*)=F_{\min}(\boldsymbol{Y}^*)$ или $Z_{\min}(\boldsymbol{X}^*)=F_{\max}(\boldsymbol{Y}^*)$.

Доказательство. Рассмотрим несимметрическую пару сопряженных задач. Возьмем математическую модель исходной задачи в виде II.1:

$$Z=\boldsymbol{C}\cdot\boldsymbol{X}\quad\to\quad\min$$
$$\boldsymbol{A}\cdot\boldsymbol{X}=\boldsymbol{A}_0 \qquad (5.1)$$
$$\boldsymbol{X}\ge\boldsymbol{0}$$

Пусть эта задача имеет решение. Тогда ее оптимальный план можно определить симплексным методом. Будем считать, что базисными векторами есть вектора $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, …, $\boldsymbol{A}_m$ и последняя итерация симплексного метода имеет вид, представленный в табл. 5.1

Таблица 5.1

Базис	$C_б$	$\boldsymbol{A}_0$	c_1 $\boldsymbol{A}_1$	c_2 $\boldsymbol{A}_2$	…	c_m $\boldsymbol{A}_m$	c_{m+1} $\boldsymbol{A}_{m+1}$	…	c_n $\boldsymbol{A}_n$
$\boldsymbol{A}_1$	c_1	x^*_1	1	0	…	0	$x_{1\,m+1}$	…	$x_{1\,n}$
$\boldsymbol{A}_2$	c_2	x^*_2	0	1	…	0	$x_{2\,m+1}$	…	$x_{2\,n}$
…	…	…	…	…	…	…	…	…	…
$\boldsymbol{A}_m$	c_m	x^*_m	0	0	…	1	$x_{m\,m+1}$	…	$x_{m\,n}$
$\Delta_j=Z_j-c_j$		Z^*_0	Z_1-c_1	Z_2-c_2	…	Z_m-c_m	$Z_{m+1}-c_{m+1}$	…	Z_n-c_n

Обозначим через $\boldsymbol{D}$ матрицу, состоящую из компонентов векторов $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, …, $\boldsymbol{A}_m$ (базисные вектора последней итерации), тогда табл. 5.1 содержит коэффициенты разложения векторов $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, …, $\boldsymbol{A}_n$ исходной системы по векторам базиса, то есть каждому вектору $\boldsymbol{A}_j$ в этой таблице соответствует вектор $\boldsymbol{X}_j$, такой что

$$\boldsymbol{A}_j=\boldsymbol{D}\cdot\boldsymbol{X}_j,\quad j=\overline{1,n}. \qquad (5.2)$$

Для оптимального плана имеем:

$$\boldsymbol{A}_0=\boldsymbol{D}\cdot\boldsymbol{X}^*,\qquad \text{де}\qquad \boldsymbol{X}^*=(x_1^*\ x_2^*\ ...\ x_n^*)^{\boldsymbol{T}}. \qquad (5.3)$$

Обозначим $\overline{\boldsymbol{X}}$ матрицу, состоящую с коэффициентов в разложении векторов $\boldsymbol{A}_j$ в табл. 5.1. С учетом соотношений (5.2)–(5.3) имеем:

$$\boldsymbol{A} = \boldsymbol{D} \cdot \overline{\boldsymbol{X}} \quad \Rightarrow \quad \boldsymbol{D}^{-1} \cdot \boldsymbol{A} = \overline{\boldsymbol{X}} \ , \tag{5.4}$$

$$\boldsymbol{A}_0 = \boldsymbol{D} \cdot \boldsymbol{X}^* \quad \Rightarrow \quad \boldsymbol{D}^{-1} \cdot \boldsymbol{A}_0 = \boldsymbol{X}^* . \tag{5.5}$$

По таблице минимальное значение целевой функции равно:

$$Z_0^* = \mathbf{C}_б \cdot \mathbf{X}^* . \tag{5.6}$$

Рассмотрим вектор $\overline{\mathbf{Z}}$:

$$\begin{aligned} \overline{\boldsymbol{Z}} &= C_б \cdot \overline{\boldsymbol{X}} - C = (C_б \cdot X_1 - c_1,\ C_б \cdot X_2 - c_2, \ldots, C_б \cdot X_n - c_n) = \\ &= (Z_1 - c_1, Z_2 - c_2, \ldots, Z_n - c_n) \end{aligned} \tag{5.7}$$

Поскольку его компоненты являются соответствующими оценками $\Delta_j = Z_j - c_j \le 0$ ($j = \overline{1,n}$) оптимального плана, то вектор $\overline{\boldsymbol{Z}}$ неположительный: $\overline{\boldsymbol{Z}} = C_б \cdot \overline{\boldsymbol{X}} - \boldsymbol{C} \le \boldsymbol{0}$.

Оптимальный план исходной задачи имеет вид: $\boldsymbol{X}^* = \boldsymbol{D}^{-1} \cdot \boldsymbol{A}_0$, поэтому оптимальный план двойственной задачи:

$$F = \boldsymbol{Y} \cdot \boldsymbol{A}_0^{\boldsymbol{T}} \quad \to \quad \max \tag{5.8}$$

$$\boldsymbol{Y} \cdot \boldsymbol{A} \le \boldsymbol{C}^{\boldsymbol{T}} .$$

Ищем в аналогичной форме: $\boldsymbol{Y}^* = \boldsymbol{C}_б \cdot \boldsymbol{D}^{-1}$. Покажем, что вектор $\boldsymbol{Y}^*$ действительно является планом двойственной задачи (5.8). Для этого ее систему ограничений запишем в виде $\boldsymbol{Y} \cdot \boldsymbol{A} - \boldsymbol{C}^{\boldsymbol{T}} \le \boldsymbol{0}$ и подставим в левую часть неравенства вектор $\boldsymbol{Y}^*$. Учитывая соотношения (5.4), (5.7) та (5.8), имеем

$$\boldsymbol{Y}^* \cdot \boldsymbol{A} - \boldsymbol{C}^{\boldsymbol{T}} = \boldsymbol{C}^* \cdot \boldsymbol{D}^{-1} \cdot \boldsymbol{A} - \boldsymbol{C}^{\boldsymbol{T}} = \boldsymbol{C}^* \boldsymbol{D}^{-1} \boldsymbol{D} \overline{\boldsymbol{X}} - \boldsymbol{C}^{\boldsymbol{T}} = \boldsymbol{C}^* \overline{\boldsymbol{X}} - \boldsymbol{C}^{\boldsymbol{T}} \le \boldsymbol{0} .$$

Таким образом, $\boldsymbol{Y}^* \cdot \boldsymbol{A} \le \boldsymbol{C}^{\boldsymbol{T}}$. Поскольку вектор $\mathbf{Y}^*$ удовлетворяет систему ограничений задачи (5.8), то $\mathbf{Y}^*$ является планом двойственной задачи. Значение целевой функции двойственной задачи с эти планом будет $F(\boldsymbol{Y}^*) = \boldsymbol{Y}^* \boldsymbol{A}_0^{\boldsymbol{T}}$. С другой стороны,

$$F(\mathbf{Y}^*) = \mathbf{Y}^* \mathbf{A}_0^{\mathbf{T}} = \mathbf{C}_б^* \cdot \mathbf{D}^{-1} \mathbf{A}_0^{\mathbf{T}} = \mathbf{C}^* \mathbf{X}^* = Z(\mathbf{X}^*) = Z_{\min} .$$

Таким образом, мы доказали, что целевая функция двойственной задачи для плана $\boldsymbol{Y}^*$ равняется минимальному значению целевой функции исходной задачи. Остается доказать, что $\boldsymbol{Y}^*$ является оптимальным планом двойственной задачи.

Умножим уравнением $\boldsymbol{AX} = \boldsymbol{A}_0$ на любой произвольный план Y двойственной задачи, а неравенство $\boldsymbol{Y} \cdot \boldsymbol{A} \leq \boldsymbol{C^T}$ умножим на любой план исходной задачи ***X***. Имеем:

$$\boldsymbol{YAX} = \boldsymbol{YA}_0 = F\boldsymbol{(Y)},$$

$$\boldsymbol{YAX} \leq \boldsymbol{C^T X} = Z\boldsymbol{(X)}.$$

Отсюда следует, что для любых произвольных планов **X** и **Y** выполняется неравенство $F(\mathbf{Y}) \leq Z(\mathbf{X})$. Это соотношение распространяется и на экстремальные значения целевой функции, т.е.

$$F_{\max}(\mathbf{Y}) \leq Z_{\min}(\mathbf{X}).$$

Из последнего неравенства следует, что целевая функция достигает максимального значения в том случае, если $F_{\max}\boldsymbol{(Y)} = Z_{\min}\boldsymbol{(X)}$, но это имеет место, если $\boldsymbol{Y} = \boldsymbol{Y}^*$, следовательно, $\boldsymbol{Y}^*$ и есть оптимальный план двойственной задачи.

Аналогично можно доказать, что если двойственная задача имеет решение, то и исходная также имеет решение, причем $Z_{\min}\boldsymbol{(X)} = F_{\max}\boldsymbol{(Y)}$.

Для доказательства второй части теоремы, предположим, что целевая функция исходной задачи не ограничена снизу. Тогда следует, что $F\boldsymbol{(Y)} \leq -\infty$. Это обозначает, что двойственная задача имеет противоречивую систему ограничений, а следовательно, не имеет решений.

Так само предположим, что целевая функция двойственной задачи не ограничена сверху, то есть $Z\boldsymbol{(X)} \geq +\infty$. Это обозначает, что исходная задача имеет неограниченную целевую функцию на множестве своих планов или ее система ограничений противоречивая, а следовательно, исходная задача нет имеет решений.

Теорема 1.2. Для произвольных решений двух задач $Z_1 \leq F_2$.

Теорема 1.3. Если в одной задаче функция цели неограниченная, то другая задача решений не имеет.

Теорема 1.4. Если одна задача не имеет решений, то в другой задаче функция цели неограниченная, или же другая задача также не имеет решения.

Теорема 2. В оптимальных решениях из двух взаимно двойственных неизвестных одна обязательно равна нулю:

$$x_j^0 \cdot v_j^0 = 0, \quad y_i^0 \cdot u_i^0 = 0.$$

В оптимальных решениях свободной неизвестной одной задачи соответствует базисная неизвестная другой задачи и наоборот.

Замечание. При решении ЗЛП симплекс-методом решение двойственной задачи располагается в строке целевой функции.

Пример. Построить и решить двойственную задачу для исходной задачи:

$$Z = x_1 - x_2 \to \max;$$

$$\begin{cases} -2x_1 + x_2 \le 2; \\ x_1 - 2x_2 \le 2; \\ x_1 + x_2 \le 5; \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

Рассмотреть решение задач с использованием теорем двойственности.

Решение.

Исходная задача:

$$Z = x_1 - x_2 \to \max;$$

$$\begin{cases} -2x_1 + x_2 \le 2; \\ x_1 - 2x_2 \le 2; \\ x_1 + x_2 \le 5; \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

Двойственная задача:

$$F = 2y_1 + 2y_2 + 5y_3 \quad \to \quad \min;$$

$$\begin{cases} -2y_1 + y_2 + y_3 \ge 1; \\ y_1 - 2y_2 + y_3 \ge -1; \\ y_i \ge 0,\ i = \overline{1,3}. \end{cases}$$

Решим исходную задачу симплексным методом, для преобразуем неравенства в равенства:

$$\begin{cases} -2x_1 + x_2 + x_3 = 2; \\ x_1 - 2x_2 + x_4 = 2; \\ x_1 + x_2 + x_5 = 5; \\ x_j \ge 0,\ j = \overline{1,5}. \end{cases}$$

Базис	$C_б$	c_j / A_0	$c_1=1$ / A_1	$c_2=-1$ / A_2	$c_3=0$ / A_3	$c_4=0$ / A_4	$c_5=0$ / A_5
A_3	0	2	-2	1	1	0	0
A_4	0	2	**1**	-2	0	1	0
A_5	0	5	1	1	0	0	1
Целевая функция		0	0	0	0	0	0
$\Delta_j = Z_j - c_j$			-1	1	0	0	0
A_3	0	6	0	-3	1	2	0
A_1	1	2	1	-2	0	1	0
A_5	0	3	0	**3**	0	-1	1

Целевая функция		2	1	-2	0	1	0
$\Delta_j = Z_j - c_j$		2	0	-1	0	1	0
A$_3$	0	9	0	0	1	1	1
A$_1$	1	4	**1**	0	0	$\frac{1}{3}$	$\frac{2}{3}$
A$_2$	-1	1	0	1	0	$-\frac{1}{3}$	$\frac{1}{3}$
Целевая функция		3	1	-1	0	$\frac{2}{3}$	$\frac{1}{3}$
$\Delta_j = Z_j - c_j$			0	0	0	$\frac{2}{3}$	$\frac{1}{3}$

Из таблицы следует, что $X^* = (4, 1)$, $Z(X^*) = 3$.

На основании 1-й теоремы двойственности имеем: $Z(X^*) = F(Y^*) = 3$.

Поскольку $x_1 \geq 0, x_2 \geq 0$, то по 2-й теореме двойственности систему ограничений двойственной задачи запишем в виде:

$$\begin{cases} -2y_1 + y_2 + y_3 = 1; \\ y_1 - 2y_2 + y_3 = -1. \end{cases}$$

Подставив $X^* = (4, 1)$ в систему ограничений исходной задачи:

$$\begin{cases} -2 \cdot 4 + 1 \leq 2; \\ 4 - 2 \cdot 1 \leq 2; \\ 4 + 1 \leq 5; \end{cases} \quad \begin{cases} -7 < 2; & \Rightarrow y_1 = 0; \\ 2 = 2; & \Rightarrow y_2 > 0; \\ 5 = 5; & \Rightarrow y_3 > 0. \end{cases}$$

Тогда система ограничений двойственной задачи примет вид:

$$\begin{cases} y_2 + y_3 = 1; \\ -2y_2 + y_3 = -1. \end{cases}$$

Откуда $Y^* = \left(0, \frac{2}{3}, \frac{1}{3}\right)$, $F(Y^*) = 3$.

Пусть известно решение двойственной задачи $Y^* = \left(0, \frac{2}{3}, \frac{1}{3}\right)$, $F(Y^*) = 3$, найдем решение исходной задачи. По 1-1 теореме двойственности $F(Y^*) = Z(X^*) = 3$. Поскольку $y_2 > 0, y_3 > 0$, то 2-й теореме двойственности второе и третье неравенства исходной задачи обращаются в равенства:

$$\begin{cases} x_1 - 2x_2 = 2; \\ x_1 + x_2 = 5. \end{cases}$$

Откуда, $X^* = (4, 1)$, $Z(X^*) = 3$.

Рассмотрим решение задач методом, основанным на взаимно однозначном соответствии между переменными: основным переменным исходной задачи соответствуют балансовые переменные двойственной, и на оборот.

Пусть двойственная задача решена симплексным методом:

$$F = 2y_1 + 2y_2 + 5y_3 \quad \to \quad \min$$

$$\begin{cases} -2y_1 + y_2 + y_3 - y_4 = 1; \\ -y_1 + 2y_2 - y_3 + y_5 = 1; \\ y_i \geq 0,\ i = \overline{1,3}. \end{cases}$$

		c_j	c_1=2	c_2=2	c_3=5	c_4=0	c_5=0
Базис	***Cб***	***A*0**	***A*1**	***A*2**	***A*3**	***A*4**	***A*5**
	0	1	-2	1	**1**	-1	0
A5	0	1	-1	2	-1	0	1
Целевая функция		0	0	0	0	0	0
$\Delta_j = Z_j - c_j$							
A3	5	1	-2	1	1	-1	0
A5	0	2	-3	**3**	0	-1	1
Целевая функция		5	-12	5	5	-5	0
$\Delta_j = Z_j - c_j$			-8	3	0	-5	0
A3	5	$\frac{1}{3}$	-1	0	1	$-\frac{2}{3}$	$-\frac{1}{3}$
A2	2	$\frac{2}{3}$	-1	1	0	$-\frac{1}{3}$	$\frac{1}{3}$
Целевая функция		3	-7	2	5	-4	-1
$\Delta_j = Z_j - c_j$			-9	0	0	-4	-1

Решение другой задачи найдем по соответствию между переменными:

Задача 1	Основные переменные $x_1 \quad x_2 \quad x_3$	Балансовые переменные $x_4 \quad x_5$
Задача 2	$y_4 \quad y_5 \quad y_1$ Балансовые переменные	$y_2 \quad y_3$ Основные переменные

Значения x_j определяем по симплексной таблице (Табл. 2) в строке Δ_i в соответствующем столбце, причем значения x_j берем по модулю:

$$x_1 \to y_4, \quad x_1 = |\Delta_4| = 4,$$

$$x_2 \to y_5, \quad x_2 = |\Delta_5| = 1.$$

Таким образом, решение исходной задачи: $X^* = (4, 1)$, $Z(X^*) = 3.$

Если исходная задача решена симплексным методом, то решение двойственной задачи может быть найдено по формуле: $Y^* = \boldsymbol{C} \cdot \boldsymbol{A}^{-1}$, где $\boldsymbol{C}$ - матрица-строка коэффициентов при базисных переменных целевой функции в оптимальном решении исходной задачи; $\boldsymbol{A}^{-1}$- обратная матрица для матрицы $\boldsymbol{A}$, являющейся матрицей коэффициентов базисных переменных системы ограничений исходной задачи в оптимальном решении. С табл. 1 имеем:

$$\boldsymbol{C} = \begin{pmatrix} & -1 & 0 \end{pmatrix},\ \boldsymbol{A} = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & 1 & 0 \end{pmatrix},\ \boldsymbol{A}^{-1} = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 1 & 1 & 1 \end{pmatrix},$$

$$Y^* = \boldsymbol{C} \cdot \boldsymbol{A}^{-1} = \begin{pmatrix} & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

Таким образом, решение двойственной задачи: $Y^* = \left(0, \frac{2}{3}, \frac{1}{3}\right), \quad F(Y^*) = 3.$

Вопросы для самопроверки:

1) Привести примеры практического использования теории двойственности.

2) Сформулировать правила составления условий взаимно двойственных задач.

3) Сформулировать основные теоремы теории двойственности.

4) Объясните экономический смысл теорем двойственности.

5) Как по решению прямой задачи найти решение двойственной?

Тема 6. Экономическая интерпретация двойственных неизвестных. Двойственный симплекс-метод

6.1. Экономическая интерпретация двойственных неизвестных

6.2. Анализ устойчивости двойственных оценок

6.3. Двойственный симплекс-метод

6.1. Экономическая интерпретация двойственных неизвестных

Обычно свободные члены ограничений типа $\sum a_{ij} x_j \le b_i$ трактуются как запасы ресурсов, коэффициенты перед неизвестными – как нормы затрат ресурсов на производство единицы продукции данного вида, балансовые неизвестные u_i - как остатки ресурсов. При этом сами неравенства показывают, что общие затраты каждого ресурса не могут превышать его запасы. Функция цели Z_1, максимум которой разыскивается, обычно трактуется как прибыль от реализации произведенной продукции. На основании теоремы 1.1. прибыль в оптимальном плане можно выразить через запасы ресурсов: $Z_{1max} = F_{2min} = \sum y_i b_i$.

Отсюда следует, что неизвестная y_i (двойственная к остатку ресурсов u_i) равна первой производной от максимальной прибыли по запасу i-го ресурса $\frac{\partial Z_{1max}}{\partial b_i} = 0$ и поэтому имеем такую интерпретацию: неизвестная показывает, на сколько увеличится прибыль в оптимальном плане, если увеличить запас ресурса b_i на единицу. Другими словами прибыль при увеличении ресурса b_i на Δb_i увеличится на $y_i \Delta b_i$, при этом вариация ресурса Δb_i не должна превышать некоторого допустимого предела.

При наличии ограничения типа $\ge$, когда свободные члены ограничений нельзя трактовать как запасы ресурсов, двойственные неизвестные интерпретируются так: двойственная неизвестная y_i показывает интенсивность увеличения функции цели в оптимальном решении при изменении свободного члена b_i на единицу в сторону ослабления соответствующего ограничения.

Таким образом, неизвестные y_i являются некими стоимостными оценками полезности ресурсов. Это не означает, что такова реальная цена ресурсов, за которую их можно приобрести; теневые цены характеризуют полезность каждого ресурса для дальнейшего увеличения прибыли при расширенном производстве.

Если какой-то ресурс имеется в избытке, остаток этого ресурса отличен от нуля $u_k > 0$, следовательно, его двойственная неизвестная равна нулю $y_i = 0$ и дальнейшее увеличение запаса этого ресурса не будет приводить к изменению прибыли (теневая цена данного ресурса равна нулю).

Если оказалось, что в оптимальном плане исчерпаны запасы сразу нескольких ресурсов (балансовые неизвестные равны нулю, соответствующие ограничения реализуются как строгие равенства), то теневые цены y_i будут показывать наиболее предпочтительный путь расширения ресурсов. Если в задаче учтены затраты на приобретение ресурсов, то наиболее выгодным будет увеличение запасов ресурсов пропорционально их теневым ценам: $\Delta b_1 : \Delta b_2 : \Delta b_3 = y_1 : y_2 : y_3$.

Экономическую интерпретацию двойственных задач и оценки дефицитности ресурсов в окрестностях оптимального плана рассмотрим на примере.

Пример. Для изготовления двух видов продукции А и В используют три вида сырья, которое предприятие может ежемесячно закупать в ограниченных объемах. Ежемесячное нахождение необходимого сырья, затраты его на изготовление единицы каждого вида продукции, а также прибыль от реализации единицы продукции представлены в табл. 6.1.

Таблица 6.1

Виды сырья	Нормы затрат		Запасы, кг
	A	B	
I	9	6	540
II	5	10	500
III	14	7	980
Прибыль, грн.	5	6	

Определить, какое количество продукции каждого вида предприятие должно выпускать ежемесячно, чтобы прибыль от ее реализации была максимальной.

Решение. Обозначим через x_1 и x_2 соответственно количество продукции А и В, которое предприятие выпускает в течении месяца. Критерием эффективности Z является прибыль от реализации продукции. Имеем математическую модель задачи:

$$Z = 5x_1 + 6x_2 \quad \to \quad \max \qquad (6.1)$$

$$\begin{cases} 9x_1 + 6x_2 \le 540 \\ 5x_1 + 10x_2 \le 500 \\ 14x_1 + 7x_2 \le 980 \\ x_1 \ge 0, \quad x_2 \ge 0 \end{cases}. \qquad (6.2)$$

Составим двойственную задачу. Пусть каждому виду сырья поставлена в соответствие оценка y_1, y_2 та y_3. Тогда общая оценка сырья F, которое используется при изготовлении продукции, составляет

$$F = 540y_1 + 500y_2 + 980y_3 \quad \to \quad \min, \qquad (6.3)$$

а система ограничений:

$$\begin{cases} 9y_1 + 5y_2 + 14y_3 \ge 5 \\ 6y_1 + 10y_2 + 7y_3 \ge 6 \\ y_1 \ge 0, \quad y_2 \ge 0, \quad y_3 \ge 0 \end{cases} . \qquad (6.4)$$

Задачи (6.1)–(6.2) та (6.3)–(6.4) образуют симметрическую пару взаимно двойственных задач. Решением исходной задачи есть оптимальный план производства продукции A и B, а двойственной – оптимальную систему оценок. Чтобы решить эти задачи, достаточно решить одну из них. Поскольку исходная задача содержит только две неизвестные, то ее можно решить графически. Затем, по теоремам двойственности решим и вторую задачу.

Решим исходную задачу симплексным методом табл. 6.2.

Таблица 6.2

			c_1=5	c_2=6	c_3=0	c_4=0	c_5=0
Базис	**Сб**	**A_0**	**A_1**	**A_2**	**A_3**	**A_4**	**A_5**
A_3		540	9	6	1	0	0
A_4		500	5	10	0	1	0
A_5		980	14	7	0	0	1
Целевая функция							
$\Delta_j = Z_j - c_j$			-5	-6	0	0	0

$$\mathbf{X}_0 = (0;\ 0;\ 540;\ 500;\ 980),\ \Delta_2 = -6,\ \theta_{22} = 50,\ \Delta_2 \cdot \theta_{22} = -300$$

A_3		240	6	0	1	-0,6	0
A_2	6	50	1/2	1	0	0,1	0
A_5		630	21/2	0	0	-0,7	1
Целевая функция		300	3	6	0	0,6	0
$\Delta_j = Z_j - c_j$			-2	0	0	0,6	0

$$\mathbf{X}_1 = (0;\ 50;\ 240;\ 0;\ 630),\ \Delta_1 = -2,\ \theta_{11} = 40,\ \Delta_1 \cdot \theta_{11} = -80$$

A_1	5	40	1	0	1/6	-0,1	0
A_2	6	30	0	1	-1/12	0,15	0
A_5		60	0	0	-7/4	0,25	1
Целевая функция		380	5	6	1/3	0,4	0
$\Delta_j = Z_j - c_j$			0	0	1/3	0,4	0

$$\mathbf{X}^* = (40;\ 30;\ 0;\ 0;\ 60), \qquad \mathbf{Y}^* = (1/3;\ 2/5;\ 0).$$

$$F_{\min} = 540 \cdot 1/3 + 500 \cdot 0{,}4 = 380.$$

Поскольку $F_{\min} = Z_{\max}$, обе задачи решены правильно.

Подставим значения двойственных оценок $y_1^*= 1/3$, $y_2^*= 2/5$ и $y_3^*= 0$ в систему ограничений двойственной задачи:

$$\begin{cases} 9 \cdot \frac{1}{3} + 5 \cdot 0{,}4 = 5 \\ 6 \cdot \frac{1}{3} + 10 \cdot 0{,}4 = 6 \end{cases}.$$

Оба ограничения двойственной задачи выполняются как уравнения, то есть двойственные оценки сырья не превышают прибыль от реализации каждого вида продукции. Поэтому продукцию целесообразно производить в количествах, которые определяются оптимальным планом исходной задачи. Любое изменение начальных условий исходной задачи может повлиять как на ее оптимальный план, так и на систему двойственных оценок, поэтому экономический анализ с использованием двойственных оценок предусматривает определение интервал их устойчивости

Переменные $y_1^*= 1/3$ и $y_2^*= 2/5$ определяют условные двойственные оценки единицы сырья соответственно I и II типов. Эти две оценки отличны от нуля и поэтому используются полностью. В это время двойственная оценка сырья III типа равна нулю $y_3^* = 0$, поэтому при оптимальном плане производства $\mathbf{X}^*$остается определенный запас сырья III типа. Таким образом, положительные условные оценки имеют только те типы сырья, которые полностью используются при оптимальном плане производства. Следовательно, двойственные оценки характеризуют дефицит используемого сырья.

Рассмотрим числа, которые содержатся в столбце вектора $\mathbf{A}_3$ последней итерации симплексной таблицы. Следовательно, рост прибыли от реализации при увеличении объема сырья I типа на единицу обусловлено переходом к новому оптимальному плану, по которому выпуск продукции А увеличивается на 1/6 единиц, выпуск продукции В уменьшается на 1/12 единиц, а использование сырья III типа возрастает на 7/4 кг. Соответственно, числа в столбце вектора $\mathbf{A}_4$, показывают, как измениться оптимальный план, если увеличить количество сырья II типа на 0,4 кг.

Величина двойственной оценки показывает, насколько возрастет максимальное значение целевой функции исходной задачи при увеличении количества сырья соответственного типа на 1кг. Так, увеличение количества сырья I типа на 1 кг приведет к тому, что появится возможность найти новый план производства, по которому ежемесячная общая прибыль от реализации изделий может увеличится на $\Delta Z = 1/3$ *грн.*

6.2. Анализ устойчивости двойственных оценок

Переменные двойственной задачи определяют темп изменения целевой функции исходной задачи, если изменяется запас i-ого вида сырья. Это справедливо в окрестности точки b_i, а именно пока базис D остается неизменным. Если $y_i^* > 0$, то значения целевой функции исходной задачи можно увеличить за счет увеличения запасов i-ого сырья, поскольку это сырье будет использоваться для увеличения выпуска продукции. Если $y_i^* = 0$, то увеличение i-ого вида сырья не сопровождается увеличением целевой функции, так как этого сырья достаточно. Приведенные соотношения соответствуют второй теореме двойственности и есть основой для теории дополняющей нежесткости.

Система уравнений (6.5) показывает, что изменение величины b_i сопровождается изменением экстремального значения целевой функции $Z_{\max}$ на величину $\left|y_i^*\right|$ и может быть охарактеризована лишь тогда, когда при изменении b_i значения переменных y_i^* в оптимальном плане не меняется. Это справедливо для тех $b_i + \Delta b_i$, для которых нет отрицательных чисел среди компонент вектора **B**, где матрица $\mathbf{D}^{-1}$ является обратной к матрице D, образованной с компонент базисных векторов, что определяли первоначальный план исходной задачи:

$$\mathbf{G} = \mathbf{D}^{-1} \cdot \begin{pmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ . \quad . \quad . \quad . \\ b_m + \Delta b_m \end{pmatrix}. \qquad (6.6)$$

Следовательно, если найдено решение исходной задачи, то без особенных трудностей можно провести анализ устойчивости двойственных оценок относительно изменения величин b_i. Это позволяет оценить устойчивость оптимального плана двойственной задачи относительно изменения b_i, определить степень влияния изменения b_i на максимальное значение целевой функции исходной задачи, что дает возможность найти наиболее целесообразное направление возможных изменений b_i.

Пример. По условию предыдущей задачи определить интервал устойчивости двойственных оценок по отношению к изменению запасов сырья каждого вида.

Решение. Найдем матрицу $\mathbf{D}^{-1}$, которая есть обратной к матрице D, образованной с компонентов базисных векторов A_3, A_4 и A_5. Матрицу $\mathbf{D}^{-1}$ можно записать по последней итерации симплексной таблицы (табл. 7.2):

$$\mathbf{D}^{-1} = \begin{pmatrix} 1/6 & -0{,}1 & 0 \\ -1/12 & 0{,}15 & 0 \\ -7/4 & 0{,}25 & 1 \end{pmatrix}.$$

Тогда компоненты вектора **G** (6.6) запишем таким образом:

$$\mathbf{D}^{-1} \cdot \begin{pmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ . \quad . \quad . \quad . \\ b_m + \Delta b_m \end{pmatrix} = \begin{pmatrix} 1/6 & -0{,}1 & 0 \\ -1/12 & 0{,}15 & 0 \\ -7/4 & 0{,}25 & 1 \end{pmatrix} \cdot \begin{pmatrix} 540 + \Delta b_1 \\ 500 + \Delta b_2. \\ 980 + \Delta b_3 \end{pmatrix} = \begin{pmatrix} 40 + 1/6\Delta b_1 - 0{,}1\Delta b_2 \\ 30 + 1/2\Delta b_1 + 0{,}15\Delta b_2 \\ 60 - 7/4\Delta b_1 + 0{,}25\Delta b_2 + \Delta b_3 \end{pmatrix} \text{O}$$

пределим, при каких значениях Δb_i $(i = \overline{1,3})$ эти компоненты неотрицательные:

$$\begin{cases} 40 + 1/6\Delta b_1 - 0{,}1\Delta b_2 \geq 0 \\ 30 - 1/2\Delta b_1 + 0{,}15\Delta b_2 \geq 0 \\ 60 - 7/4\Delta b_1 + 0{,}25\Delta b_2 + \Delta b_3 \geq 0 \end{cases} . \qquad (6.7)$$

Если запасы сырья I и II типов не меняются, система имеет решение:

$$\begin{cases} \Delta b_1 = 0 \\ \Delta b_2 = 0 \\ \Delta b_3 \geq -60 \end{cases} \Rightarrow \quad \Delta b_3 \in [-60; +\infty) \; .$$

То есть запасы сырья III типа могут неограниченно возрастать или уменьшаться в пределах 60 кг, что не сопровождается изменением плана двойственной задачи: $\mathbf{Y}^* = (/3;\ 2/5;\ 0)$. Поскольку по этому плану оценка, соответствующая сырью III типа, равна нулю, экстремальное значение целевой функции при изменении количества сырья III типа не меняется.

Если остаются неизменными запасы сырья I и III типов, система неравенств (6.7) дает решение:

$$\begin{cases} \Delta b_1 = 0 \\ \Delta b_3 = 0 \\ -240 \le \Delta b_2 \le 400 \end{cases} \quad \Rightarrow \quad \Delta b_2 \in [-240;400] \ .$$

При изменении сырья II типа в пределах $\Delta b_2 \in [-240;400]$ экстремальное значение целевой функции изменяется на величину $\Delta Z_{\max} = {}^{2}\!/_{5} \cdot \Delta b_2$, что связано с изменением оптимального плана производства.

При постоянных запасах сырья II и III типов изменение запасов сырья I типа в пределах $\Delta b_1 \in [-240;\, 240/7]$ не приведет к изменению оптимального плану двойственной задачи, а экстремальное значение целевой функции (при соответствующему изменению плану производства) изменится на величину $\Delta Z_{\max} = {}^{1}\!/_{3} \cdot \Delta b_1$.

Следовательно, если меняются запасы сырья только одного из типов, интервал устойчивости двойственных оценок следующий:

$$\Delta b_3 \in [-60;+\infty), \qquad \text{если } \Delta b_1 = 0, \quad \Delta b_2 = 0;$$

$$\Delta b_2 \in [-240;400], \qquad \text{если } \Delta b_1 = 0, \quad \Delta b_3 = 0;$$

$$\Delta b_1 \in [-240;\, 240/7], \text{ если } \Delta b_2 = 0, \quad \Delta b_3 = 0.$$

Проверим, меняется ли оптимальный план двойственной задачи, если увеличить запасы всех видов сырья на 20, 30 и 10 кг соответственно. Для этого подставим данные, приведенные в системе неравенств (6.7):

$$\begin{cases} 40 + 1/6 \cdot 20 - 0{,}1 \cdot 30 > 0 \\ 30 - 1/2 \cdot 20 + 0{,}15 \cdot 30 > 0 \\ 60 - 7/4 \cdot 20 + 0{,}25 \cdot 30 + 10 > 0 \end{cases} .$$

Поскольку система неравенств выполняется, то оптимальный план двойственной задачи остается неизменным. Увеличение сырья приведет к увеличению максимального значения целевой функции исходной задачи на величину:

$$\Delta Z_{\max} = \Delta b_1 \cdot y_1^* + \Delta b_2 \cdot y_2^* + \Delta b_3 \cdot y_3^* = 20 \cdot {}^{1}\!/_{3} + 30 \cdot {}^{2}\!/_{5} + 10 \cdot 0 = {}^{56}\!/_{3} .$$

Следует заметить, что целесообразно рассмотреть экономическое содержание требований положительности оценок исходной задачи, если ее целевая функция исследуется на максимум. Если $\Delta_j = Z_j - c_j < 0$, то удельный вес стоимости ресурсов Z_j, которые необходимы для увеличения уровня производственной деятельности на единицу, меньше чем удельная прибыль, поэтому необходимо ввести x_j в базис, что эквивалентно увеличению небазисной переменной x_j от нуля до определенного положительного значения. Следовательно, переходим к новому опорному плану. Процесс продолжается до тех пор, пока все оценки станут положительными, то есть для всех видов сырья удельный вес станет равным или меньшим чем удельная прибыль.

6.3. Двойственный симплекс-метод

В теории двойственности исходная задача решается симплекс-методом, двойственная задача решается двойственным симплекс-методом. Существует целый класс задач, например задача о диете, условия которой естественным образом формулируются сразу во второй стандартной форме. Для решения таких задач двойственный симплекс-метод более удобен, чем обычный алгоритм симплекс-метода.

Рассмотрим двойственный симплекс-метод на примере.

Пример. Найти $F = 2x_1 + 3x_2 \to \min$ при ограничениях:

$$\begin{cases} 3x_1 + 2x_2 \ge 6 \\ x_1 + 4x_2 \ge 4 \\ x_1 \ge 0,\ x_2 \ge 0 \end{cases}$$

Решение:

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

№	Базис	$c_{баз}$	c_j	2	3	0	0	Примечания
			A_0	A_1	A_2	A_3	A_4	
1	A_3	0	-6	-3	-2	1	0	$\theta = \min\left\{\frac{-b_i}{-a_{ik}}\right\}$
2	A_4	0	-4	-1	-4	0	1	
	Δ_1		0	-2	-3	0	0	
3	A_1	2	2	1	2/3	-1/3	0	(3)=(1):[-3]
4	A_4	0	-2	0	-10/3	-1/3	1	(4)=(2)+(3)
	Δ_2		4	0	-5/3	-2/3	0	
5	A_1	2	8/5	1	0	-2/5	1/5	(5)=(3)+(6)[-2/3]
6	A_2	3	3/5	0	1	1/10	-3/10	(6)=(4):[-10/3]
	Δ_3		5	0	0	-1/2	-1/2	

$F_{\min} = Z_{\max} = 5 \quad X^* = (8/5, 3/5, 0, 0) \quad Y^* = (1/2, 1/2, 0, 0)$

Вопросы для самопроверки:

1) Дайте экономическую интерпретацию свойств двойственных оценок.
2) Какая экономическая интерпретация двойственных неизвестных?
3) В чем заключается анализ устойчивости двойственных оценок?
4) В чем суть двойственного симплекс-метода?

Тема 7. Анализ линейных моделей экономических оптимизационных задач

7.1. Задачи ЛП с параметрами в свободных членах ограничений

Знание двойственных неизвестных y_i позволяет установить относительную ценность каждого дефицитного ресурса и выяснить, запасы каких ресурсов выгодно изменять, и в какой пропорции. Однако неясно, до каких пределов можно увеличивать запас одного, пусть даже самого дефицитного ресурса, т.к. с его увеличением будут исчерпаны запасы остальных ресурсов и уже недостача других ресурсов будет сдерживать возможный дальнейший рост прибыли. Для того чтобы выяснить, в каких пределах вариации ресурсов найденный план остается оптимальным и допустимым, требуется решать задачу с параметрами. При решении задач ЛП с параметрами в свободных членах $b = b_0 + b_1 \cdot t_1 + b_2 \cdot t_2$ в симплекс–таблицу вводим несколько столбцов свободных членов b_0, t_1, t_2. После определения оптимального решения по первому (основному) столбцу b_0 следует выписать комбинированное решение с учетом остальных столбцов свободных членов и произвести анализ зависимости решения от параметров. Если при изменении параметров комбинированный свободный член в каком–либо ограничении изменяет знак и становится отрицательным, то полученное решение становится недопустимым. В этом случае удобно воспользоваться итерацией двойственного симплекс–метода, в результате которой условие оптимальности не нарушится и будет получено оптимальное решение для другой области изменения параметра.

Учет параметров необходим для проверки устойчивости найденного статичного решения, для определения его чувствительности к неизбежным малым нарушениям условий задачи. Для иллюстрации этого рассмотрим задачу.

Пример. В таблице приведены нормы расхода ресурсов (сталь, цветные металлы, загрузка токарных станков) для производства двух видов изделий А и В. В этой же таблице указана прибыль от производства и реализации каждого изделия.

Виды ресурсов	A	B	Запасы
Сталь, кг	10	70	570
Цветные металлы, кг	20	50	420
Токарные станки, станко–ч	300	400	5600
Прибыль, тыс. грн	3	8	

Требуется найти оптимальный план производства продукции и проверить его на устойчивость по отношению к вариациям запасов ресурсов.

Обозначим через x_1, x_2 – количество изделий каждого вида и сформулируем условия задачи в виде:

Требуется найти максимальное значение функции цели

$$z = 3x_1 + 8x_2 \rightarrow max$$

при ограничениях:

$$\begin{cases} 10x_1 + \ \ 70x_2 \leq \ \ 570 \\ 20x_1 + \ \ 50x_2 \leq \ \ 420 \\ 300x_1 + 400x_2 \leq 5600 \end{cases}$$

$$\text{все } \ x_j \geq 0.$$

Можно сократить первые два неравенства на 10, а третье неравенство – на 100; это приведет к тому, что нормы расхода a_{ij}, запасы b_i, остатки y_i и вариации $t_i = \Delta b_i$ ресурсов будут измеряться в десятках кг и сотнях станко–часов соответственно.

В этой задаче необходимо найти не только одно статичное решение для конкретных значений свободных членов, но также исследовать оптимальное решение на устойчивость при возможных вариациях свободных членов. Экономическая интерпретация – требуется не только получить оптимальный план производства продукции при заданных запасах ресурсов, но также разрешить некоторые производственные вопросы, например: какие ресурсы можно сократить и на сколько, какие ресурсы выгодно увеличить, в какой пропорции и на сколько. Поэтому общий столбец свободных членов представляем в виде четырех слагаемых

$b = b_0 + b_1\ t_1 + b_2\ t_2 + b_3\ t_3$ и вводим 3 параметра – по одному на каждый свободный член. Здесь b_0 – исходный столбец свободных членов, а коэффициенты столбцов параметров t_1, t_2, t_3 образуют в симплекс–таблице единичную матрицу.

Переписываем систему ограничений с учетом параметров

$$\begin{cases} x_1 + 7x_2 \le 57 + t_1 \\ 2x_1 + 5x_2 \le 42 + t_2 \\ 3x_1 + 4x_2 \le 56 + t_3 \end{cases},$$

приводим задачу к канонической форме и заполняем симплекс–таблицу:

		c_j				3	8	0	0	0	
№	**Б**	b_0	t_1	t_2	t_3	x_1	x_2	x_3	x_4	x_5	*Примечания*
1	x_3	57	1			1	7	1			
2	x_4	42		1		2	5		1		
3	x_5	56			1	3	4			1	
4	Δ_1	$Z_1 = 0$				-3	-8				

Решаем задачу, ориентируясь только на основной столбец свободных членов b_0, не обращая пока внимания на столбцы параметров t_1, t_2, t_3. Сначала вводим в базис x_2.

		c_j				3	8	0	0	0	
№	**Б**	b_0	t_1	t_2	t_3	x_1	x_2	x_3	x_4	x_5	*Примечания*
5	x_2	$^{57}/_7$	$^1/_7$	0	0	$^1/_7$	1	$^1/_7$	0	0	[1] : 7
6	x_4	$^9/_7$	$-^5/_7$	1	0	$^9/_7$	0	$-^5/_7$	1	0	[2] – [5]×5
7	x_5	$^{164}/_7$	$-^4/_7$	0	1	$^{17}/_7$	0	$-^4/_7$	0	1	[3] – [5]×4
8	Δ_2	$Z_2 = {}^{456}/_7$	$^8/_7$	0	0	$-^{13}/_7$	0	$^8/_7$	0	0	

Теперь вводим в базис x_1, т.к. $\Delta_2 < 0$

		c_j				3	8	0	0	0	
№	**Б**	b_0	t_1	t_2	t_3	x_1	x_2	x_3	x_4	x_5	*Примечания*
9	x_1	1	$-^5/_9$	$^7/_9$	0	1	0	$-^5/_9$	$^7/_9$	0	[6] : $^9/_7$
10	x_2	8	$^2/_9$	$-^1/_9$	0	0	1	$^2/_9$	$-^1/_9$	0	[5] – [9]×$^1/_7$
11	x_5	21	$^7/_9$	$-^{17}/_9$	1	0	0	$^7/_9$	$-^{17}/_9$	1	[7] – [9]×$^{17}/_7$
12	Δ_3	$Z_3 = 67$	$^1/_9$	$^{13}/_9$	0	0	0	$^1/_9$	$^{13}/_9$	0	

После этих двух итераций в строке 12 (строке функции цели) больше нет отрицательных элементов. Выписываем полученное оптимальное решение вместе с информацией в столбцах t_1, t_2, t_3:

$$x_1 = 1 - {}^5/_9\, t_1 + {}^7/_9\, t_2;$$

$$x_2 = 8 + {}^2/_9\, t_1 - {}^1/_9\, t_2;$$

$$x_5 = 21 + {}^7/_9\, t_1 - {}^{17}/_9\, t_2 + t_3;$$

$$x_3 = x_4 = 0;$$

$$z_{max} = 67 + {}^{1}\!/_{9}\, t_1 + {}^{13}\!/_{9}\, t_2 .$$

Напомним, что свободные члены функции цели записаны в симплекс–таблице с обратными знаками.

Заметим, что столбцы параметров $t_1 = \Delta b_1$, $t_2 = \Delta b_2$, $t_3 = \Delta b_3$ в таблицах полностью дублируются столбцами остатков ресурсов x_3 , x_4, x_5, поэтому не было особой необходимости их специально вычислять. В данном примере все ограничения были ограничениями–неравенствами одного типа (типа $\le$), поэтому столбцы при балансовых неизвестных в исходной таблице полностью совпадали со столбцами параметров. При ограничениях–неравенствах противоположного типа (типа $\ge$) столбцы параметров отличаются от столбцов балансовых неизвестных только знаком. Во всяком случае, при решении большинства задач ЛП из стандартной симплекс–таблицы можно выписать не одно статичное решение, а более общее решение с учетом возможных вариаций запасов ресурсов (столбцы параметров требуется вводить только в случае ограничений–равенств, для которых балансовые неизвестные тождественно равны нулю).

В данной задаче получилось, что прибыль в оптимальном плане не зависит от $t_3 = \Delta b_3$. Это естественно, т.к. в этом оптимальном плане третий ресурс не израсходован полностью, остаток этого ресурса отличен от нуля $x_5 = 21$ (т.е. можно сократить избыточный запас $b_3 = 56$ на 21 единицу без сокращения производства). Первые два ресурса израсходованы полностью $x_3 = x_4 = 0$, и именно это сдерживает дальнейший рост прибыли. Рассматривая зависимость прибыли от параметров $z_{max} = 67 + {}^{1}\!/_{9}\, t_1 + {}^{13}\!/_{9}\, t_2$, замечаем, что увеличение запаса 1-го ресурса на единицу (на 10 кг) приводит к увеличению прибыли на ${}^{1}\!/_{9}$, а увеличение запаса 2-го ресурса приводит к увеличению прибыли на ${}^{13}\!/_{9}$ стоимостных единиц (тыс. грн.). Из этих первых двух ресурсов второй является более дефицитным, более полезным для получения максимальной прибыли. При расширении производства надо в первую очередь вкладывать средства в увеличение запаса второго ресурса.

Наибольшая отдача будет при увеличении ресурсов в пропорции

$\Delta b_1 : \Delta b_2 : \Delta b_3 = {}^{1}\!/_{9} : {}^{13}\!/_{9} : 0 = 1 : 13 : 0$.

Но как узнать, до каких пределов можно изменять запасы ресурсов ?

Будем сначала варьировать запас самого дефицитного второго ресурса (цветные металлы) и рассмотрим, в каких пределах изменения параметра t_2

найденный план остается допустимым. При любых возможных значениях параметра компоненты планы должны оставаться неотрицательными:

$$x_1 = 1 + {}^{7}\!/_{9}\, t_2 \geq 0; \qquad \text{откуда} \quad t_2 \geq -{}^{9}\!/_{7} = -1{,}286;$$

$$x_2 = 8 - {}^{1}\!/_{9}\, t_2 \geq 0; \qquad \text{откуда} \quad t_2 \leq 72\,;$$

$$x_5 = 21 - {}^{17}\!/_{9}\, t_2 \geq 0; \qquad \text{откуда} \quad t_2 \leq {}^{189}\!/_{17} = 11{,}118;$$

$$\boldsymbol{x_3} = \boldsymbol{x_4} = 0;$$

$$z_{max} = 67 + {}^{13}\!/_{9}\, t_2\,.$$

Оказывается, найденный план остается оптимальным и допустимым только в интервале варьирования параметра $-1{,}286 \leq t_2 \leq 11{,}118$, или же для таких значений запаса второго ресурса: $40{,}714 \leq b_2 \leq 53{,}118$. Озабоченность здесь вызывает нижняя граница – стоит только уменьшить заданный запас $b_2 = 43$ всего на 1,3 единицы (т.е. на 3%), как план становится недопустимым (компонента x_1 будет отрицательной). Как говорят, найденный план неустойчив по отношению к малым изменениям свободных членов. Естественно, чтобы не было неприятной неожиданности, надо заранее найти оптимальное решение для $b_2 \leq 40{,}714$.

Продублируем ниже предыдущую таблицу (строки 9 – 12):

№	Б	b_0	t_1	t_2	t_3	x_1	x_2	x_3	x_4	x_5	*Примечания*
9	x_1	1	$-{}^{5}\!/_{9}$	${}^{7}\!/_{9}$	0	1	0	$\boxed{-{}^{5}\!/_{9}}$	${}^{7}\!/_{9}$	0	[1] : ${}^{9}\!/_{7}$
10	x_2	8	${}^{2}\!/_{9}$	$-{}^{1}\!/_{9}$	0	0	1	${}^{2}\!/_{9}$	$-{}^{1}\!/_{9}$	0	[5] – [9]×${}^{1}\!/_{7}$
11	x_5	21	${}^{7}\!/_{9}$	$-{}^{17}\!/_{9}$	1	0	0	${}^{7}\!/_{9}$	$-{}^{17}\!/_{9}$	1	[7] – [9]×${}^{17}\!/_{7}$
12	Δ_3	$Z_3 =$ 67	${}^{1}\!/_{9}$	${}^{13}\!/_{9}$	*0*	*0*	*0*	${}^{1}\!/_{9}$	${}^{13}\!/_{9}$	*0*	

При $b_2 < 40{,}714$ ($t_2 < -1{,}286$) комбинированный свободный член в 1–м ограничении (9–я строка таблицы) $b_2 = 1 + {}^{7}\!/_{9}\, t_2$ становится отрицательным.

Исправляем эту ситуацию одной итерацией двойственного симплекс–метода, для чего в строке 9 находим отрицательные элементы (если таковых нет, задача для этих значений параметра решения не имеет); в строке 9 имеется только один отрицательный элемент (выделен рамочкой и цветом), берем его за разрешающий и производим замену базиса итерацией Гаусса – Жордана:

№	Б	b_0	t_1	t_2	t_3	x_1	x_2	x_3	x_4	x_5	*Примечания*
13	x_3	$-{}^{9}\!/_{5}$	1	$-{}^{7}\!/_{5}$	*0*	$-{}^{9}\!/_{5}$	0	1	$-{}^{7}\!/_{5}$	0	[9] : ($-{}^{5}\!/_{9}$)
14	x_2	${}^{42}\!/_{5}$	0	${}^{1}\!/_{5}$	*0*	${}^{2}\!/_{5}$	1	0	${}^{1}\!/_{5}$	0	[10] – [13]×${}^{1}\!/_{7}$
15	x_5	${}^{112}\!/_{5}$	0	$-{}^{4}\!/_{5}$	*1*	${}^{7}\!/_{5}$	0	0	$-{}^{4}\!/_{5}$	1	[11] – [13]×${}^{17}\!/_{7}$
16	Δ	$Z = {}^{336}\!/_{5}$	0	${}^{8}\!/_{5}$	*0*	${}^{1}\!/_{5}$	0	0	${}^{8}\!/_{5}$	0	

Выписываем новое оптимальное решение и определяем допустимые пределы варьирования параметра t_2, для которых справедливо найденное решение:

$$x_3 = -{}^9\!/_5 - {}^7\!/_5\, t_2 \ge 0; \qquad \text{откуда} \quad t_2 \le -{}^9\!/_7 = -1{,}286;$$
$$x_2 = {}^{42}\!/_5 + {}^1\!/_5\, t_2 \ge 0; \qquad \text{откуда} \quad t_2 \ge -42\,;$$
$$x_5 = {}^{112}\!/_5 - {}^4\!/_5\, t_2 \ge 0; \qquad \text{откуда} \quad t_2 \le 28;$$
$$x_1 = x_4 = 0;$$
$$z_{max} = {}^{336}\!/_5 + {}^8\!/_5\, t_2\,.$$

Новое решение справедливо на интервале $-42 \le t_2 \le -1{,}286$, или $0 \le b_2 \le 40{,}714$. Из физических соображений очевидно, что параметр t_2 не может быть меньше -42, т.к. запас второго ресурса не может быть отрицательным (кстати, из строки 14 следует, что при $t_2 < -42$ неотрицательных решений нет, т.к. в этой строке с отрицательным свободным членом нет отрицательных коэффициентов при свободных неизвестных).

Исключительно в учебных целях найдем оптимальные решения данной задачи для всех значений параметра t_2 ($-42 \le t_2 < \infty$) при $t_1 = t_3 = 0$. Уже найдены решения в диапазонах $-42 \le t_2 \le -1{,}286$ и $-1{,}286 \le t_2 \le 11{,}118$. Для анализа диапазона $t_2 > 11{,}118$ надо вернуться к предыдущей таблице (строки 9–12).

№	Б	b_0	t_1	t_2	t_3	x_1	x_2	x_3	x_4	x_5	*Примечания*
9	x_1	1	$-{}^5\!/_9$	${}^7\!/_9$	0	1	0	$-{}^5\!/_9$	${}^7\!/_9$	0	
10	x_2	8	${}^2\!/_9$	$-{}^1\!/_9$	0	0	1	${}^2\!/_9$	$-{}^1\!/_9$	0	
11	x_5	21	${}^7\!/_9$	$-{}^{17}\!/_9$	1	0	0	${}^7\!/_9$	$-{}^{17}\!/_9$	1	
12	Δ	Z = 67	${}^1\!/_9$	${}^{13}\!/_9$	0	0	0	${}^1\!/_9$	${}^{13}\!/_9$	0	

Снова анализируя решение на интервале $-1{,}286 \le t_2 \le 11{,}118$, убеждаемся, что при $t_2 > 11{,}118$ комбинированный свободный член $b_3 = 21 - {}^{17}\!/_9\, t_2$ в 3–м ограничении (строка 11) меняет знак. В этой строке имеется один отрицательный элемент, принимаем его за разрешающий; в результате итерации двойственного симплекс–метода с этим разрешающим элементом получаем (в строках 17–20) очередное решение, которое справедливо для всех $t_2 > 11{,}118$:

№	Б	b_0	t_1	t_2	t_3	x_1	x_2	x_3	x_4	x_5	*Примечания*
17	x_4	$-{}^{189}\!/_{17}$	$-{}^7\!/_{17}$	1	$-{}^9\!/_{17}$	0	0	$-{}^7\!/_{17}$	1	$-{}^9\!/_{17}$	[11] : ($-{}^{17}\!/_9$)
18	x_1	${}^{164}\!/_{17}$	$-{}^4\!/_{17}$	0	${}^7\!/_{17}$	1	0	$-{}^4\!/_{17}$	0	${}^7\!/_{17}$	[9] – [17]×${}^7\!/_9$
19	x_2	${}^{115}\!/_{17}$	${}^3\!/_{17}$	0	$-{}^1\!/_{17}$	0	1	${}^3\!/_{17}$	0	$-{}^1\!/_{17}$	[10] + [17]×${}^1\!/_9$
20	Δ	Z =	${}^{12}\!/_{17}$	0	${}^{13}\!/_{17}$	0	0	${}^{12}\!/_{17}$	0	${}^{13}\!/_{17}$	

		$^{1412}/_{17}$									

Выписываем последнее решение:

$$x_4 = -{}^{189}/_{17} + t_2 \geq 0; \qquad \text{откуда} \quad t_2 \geq {}^{189}/_{17} = 11{,}118;$$
$$x_1 = {}^{164}/_{17} \geq 0;$$
$$x_2 = {}^{115}/_{17} \geq 0;$$
$$z_{max} = {}^{1412}/_{17}.$$

Итак, найдены оптимальные решения для всех значений параметра t_2. Ниже приведена сводная таблица решений для всех значений запаса цветных металлов $b_2 = 42 + t_2$, десятки кг:

$b_2 < 0$	Решений нет		Нет
$0 \leq b_2 \leq 40{,}714$	$x_1 = 0$	$x_2 = 0{,}2\ b_2$	$z_{max} = 1{,}6\ b_2$
$40{,}714 \leq b_2 \leq 53{,}118$	$x_1 = 0{,}778\ b_2 - 31{,}667$	$x_2 = 12{,}667 - 0{,}111\ b_2$	$z_{max} = 1{,}444\ b_2 + 6{,}333$
$b_2 > 53{,}118$	$x_1 = 9{,}647$	$x_2 = 8{,}765$	$z_{max} = 83{,}059$

Еще раз отметим, что столбцы параметров t_i полностью дублируются в таблицах столбцами балансовых неизвестных y_i, поэтому не было особой необходимости заранее предусматривать их в исходной симплекс–таблице.

Вопросы для самопроверки:

1) Сформулировать задачу линейного программирования с параметрами в свободных членах.
2) Для каких целей нужен учет параметров в свободных членах?
3) Привести этапы решения задачи линейного программирования с параметрами в свободных членах.

Тема 8. Транспортная задача. Методы решения транспортной задачи

8.1. Общая постановка транспортной задачи.

8.2Способы составления первого базисного плана. Критерий оптимальности. Метод потенциалов.

8.3. Вырождение плана транспортной задачи.

8.1. Общая постановка транспортной задачи.

Имеем m пунктов отправления $A_1, A_2, ..., A_m$ и n пунктов назначения $B_1, B_2, ..., B_n$. Известны запасы a_i продукции у каждого поставщика A_i, потребности b_j каждого потребителя B_j и тарифы c_{ij} (стоимость перевозки единицы товара от пункта A_i в пункт B_j). Требуется составить оптимальный план перевозки, т.е. определить, какое количество груза x_{ij} должно быть отправлено из каждого пункта отправления в каждый пункт назначения, чтобы общая стоимость всех перевозок была наименьшей.

Дано: тарифы, запасы и потребности

Постав-щики	Потребители					Запа-сы
	B_1	B_2	B_3	B_4	B_5	
A_1	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	a_1
A_2	c_{21}	c_{22}	c_{23}	c_{24}	c_{25}	a_2
A_3	c_{31}	c_{32}	c_{33}	c_{34}	c_{35}	a_3
Потреб-ности	b_1	b_2	b_3	b_4	b_5	

Найти: оптимальный план перевозки грузов

Постав-щики	Потребители					Запа-сы
	B_1	B_2	B_3	B_4	B_5	
A_1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	a_1
A_2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	a_2
A_3	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	a_3
Потреб-ности	b_1	b_2	b_3	b_4	b_5	

Математическая модель транспортной задачи:

Общие расходы на перевозку товара должна быть наименьшими:

$$F = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} x_{ij} \to \min$$

количество вывезенного товара от A_i не может превышать его запаса a_i:

$$\sum_{j=1}^{n} x_{ij} \le a_i ;$$

количество привезенного товара в B_j не может превышать потребности b_j:

$$\sum_{i=1}^{m} x_{ij} \le b_j .$$

Если $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, транспортная задача называется сбалансированной или закрытой. В этом случае весь произведенный товар должен быть вывезен, все потребности должны быть удовлетворены:

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i \\ \sum_{i=1}^{m} x_{ij} = b_j \end{cases}.$$

Если транспортная задача несбалансированная, то ее можно закрыть введением фиктивного поставщика или фиктивного потребителя:

$\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$ вводится фиктивный поставщик $A_ф$

$a_ф = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i$, тарифы при этом $c_ф = 0$;

$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$ вводится фиктивный потребитель $B_ф$

$b_ф = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$, тарифы при этом $c_ф = 0$.

В общей задаче линейного программирования могут быть ситуации, когда задача не имеет решения или ее функция цели неограниченная. Этого никогда не может быть для закрытой транспортной задачи. Функция цели транспортной задачи, минимум которой разыскивается, состоит из суммы неотрицательных слагаемых и ограничена снизу $F_{\min} = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} x_{ij} > 0$, поэтому она никак не может быть неограниченной.

8.2. Способы составления первого базисного плана. Критерий оптимальности. Метод потенциалов.

Оптимальный план перевозок надо разыскивать только среди базисных планов, в которых число открытых маршрутов (ненулевых x_{ij}) наименьшее. Система ограничений имеет $m+n$ равенств, но так как $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, то одно из уравнений оказывается лишним, т.е. $r = m + n - 1$, но иногда независимых ограничений еще меньше, тогда имеем случай вырождения.

Для составления первого базисного плана разработано несколько методов, из которых самыми простыми являются диагональный метод (северо-западного угла) и метод наименьшей стоимости.

Пример.

Постав-щики	Потребители				Запа-сы
	B_1	B_2	B_3	B_4	
A_1	4	3	2	7	46
A_2	1	1	6	4	34
A_3	3	5	9	4	40
Потреб-ности	40	**35**	30	45	120 150

Диагональный метод:

Постав-щики	Потребители				Запа-сы
	B_1	B_2	B_3	B_4	
A_1	4 40	3 6	2 0	7 0	46/6/0
A_2	1 0	1 29	6 5	4 0	34/5/0
A_3	3	5	9	4	40/15

	0	0	25	15	
A_4	0 0	0 0	0 0	0 30	30
Потреб-ности	40	**35/29/0**	30/25	45	150 150

$$F = 40 \cdot 4 + 6 \cdot 3 + 29 \cdot 1 + 5 \cdot 6 + 25 \cdot 9 + 15 \cdot 4 = 522 \text{ грн.}$$

Метод минимальной стоимости:

Постав-щики	Потребители				Запа-сы
	B_1	B_2	B_3	B_4	
A_1	4 0	3 16	2 30	7 0	46/16
A_2	1 34	1 0	6 0	4 0	34/0
A_3	3 6	5 0	9 0	4 34	40/34
A_4	0 0	0 19	0 0	0 11	30
Потреб-ности	40/6	**35/19**	30	45	150 150

$$F = 3 \cdot 16 + 2 \cdot 30 + 1 \cdot 34 + 6 \cdot 3 + 4 \cdot 34 = 296 \text{ грн}$$

Критерий оптимальности. Метод потенциалов

Найденное исходное опорное решение проверяется на оптимальность методом потенциалов по критерию, который утверждает следующая теорема.

Теорема. Если план $X^*\left(x_{ij}^*\right)$ транспортной задачи является оптимальным, то ему соответствует система из $m+n$ чисел u_i^* и v_j^*, удовлетворяющих условиям:

$u_i^* + v_j^* = c_{ij}$ для $x_{ij}^* > 0,$

$u_i^* + v_j^* \le c_{ij}$ для $x_{ij}^* = 0.$

Числа u_i^* и v_j^* называются потенциалами, соответственно, поставщиков и потребителей.

Потенциалы u_i и v_j находят из равенства $u_i + v_j = c_{ij}$, справедливого для занятых клеток. Одному из потенциалов дается произвольное значение, например $u_1 = 0$, тогда остальные потенциалы определяются однозначно.

На основе теоремы: для того, чтобы первоначальный опорный план был оптимальный, необходимо выполнение следующих условий:

1) для каждой занятой клетки сумма потенциалов должна быть равна стоимости единицы перевозки, стоящей в этой клетке:

$$u_i + v_j = c_{ij};$$

2) для каждой незанятой клетки сумма потенциалов должна быть меньше или равна стоимости единицы перевозки, стоящей в этой клетке:

$$u_i + v_j \le c_{ij}.$$

Если хотя бы одна незанятая клетка не удовлетворяет условию, то опорный план является неоптимальным и его можно улучшить, вводя в базис вектор, соответствующий клетке, для которой нарушается условие оптимальности.

Таким образом, для проверки плана на оптимальность необходимо сначала построить систему потенциалов.

Часто вычисляют $\Delta_{ij} = (u_i + v_j) - c_{ij}$. Эту оценку называют оценкой свободных клеток. Если $\Delta_{ij} \le 0$, то опорное решение является оптимальным. Если хотя бы одна из оценок $\Delta_{ij} > 0$, то опорное решение не является оптимальным и его можно улучшить, перейдя от одного опорного решения к другому.

Продолжение примера

	$v_1 = 2$	$v_2 = 3$	$v_3 = 2$	$v_4 = 3$
$u_1 = 0$	2 / 4	**3**	2	3 / 7
$u_2 = -1$	**1**	2 / 1	1 / 6	2 / 4
$u_3 = 1$	3	4 / 5	3 / 9	4
$u_4 = -3$	-1 / 0	0	-1 / 0	0

Для вычисления неизвестных u_i и v_j используем уравнения:

$u_1 + v_2 = 3, \quad u_1 = 0,$

$u_1 + v_3 = 2,$

$u_2 + v_1 = 1,$

$u_3 + v_1 = 3,$

$u_3 + v_4 = 4,$

$u_4 + v_2 = 0,$

$u_4 + v_4 = 0.$

Вычисляем оценки Δ_{ij}. Видим, что $\Delta_{22} = (-1 + 3) - 1 > 0$. Следовательно, опорный план не является оптимальным и его следует улучшить.

Переход к другому базисному плану

Наличие положительной оценки свободной клетки ($\Delta_{ij} > 0$) при проверке опорного решения на оптимальность свидетельствует о том, что полученное решение не оптимально и уменьшения значения целевой функции надо перейти к другому опорному решению. При этом надо перераспределить грузы, перемещая их из занятых клеток в свободные. Свободная клетка становится занятой, а одна из ранее занятых клеток – свободной.

Для свободной клетки с $\Delta_{ij} > 0$ строится цикл (цепь), все вершины которого кроме одной находятся в занятых клетках; углы прямые, число вершин четное. Около свободной клетки цикла с $\Delta_{ij} > 0$ ставится знак «+», затем поочередно проставляются знаки «-» и «+». У вершин со знаком «-» выбирают минимальный груз (ρ), его прибавляют к грузам, стоящим у вершин со знаком «+», и отнимают от грузов у вершин со знаком «-». В результате перераспределения груза получим новое опорное решение. Это решение проверяем на оптимальность, и так далее до тех пор, пока не получим оптимальное решение.

Оптимальный план должен разыскиваться только среди базисных планов с одним и тем же числом открытых маршрутов (число ненулевых базисных неизвестных неизменно и равно рангу матрицы системы ограничений задачи). Если открывается новый маршрут (в базис вводиться неизвестная x_{pq}), один из старых маршрутов должен быть закрыт.

При назначении новых объемов перевозок x_{ij} надо следить за соблюдением балансов по строкам и столбцам таблицы (все запасы должны быть вывезены, все потребности – удовлетворены).

Продолжение примера

Предположим, что мы решили изменить план, составленный по методу минимальной стоимости, открыв новый маршрут от A_2 к B_2 с тарифом $c_{22} = 1$ (поскольку $\Delta_{22} > 0$). Вводим в базис неизвестную x_{22}.

Постав- щики	Потребители			
	B_1	B_2	B_3	B_4
A_1	4 0	3 16	2 30	7 0
A_2	1 34 -	1 $+\rho$	6 0	0
A_3	3 6 +	5 0	9 0	4 34 –
A_4	0 0	0 19 –	0 0	0 11 +

Направление обхода цепи не имеет значения. Важно, чтобы все узлы цепи, кроме начального, располагались в занятых базисных клетках.

Общее число маршрутов не должно увеличиться (в противном случае новый план не будет базисным). Выбираем среди клеток со знаком «-» груз минимальной величины: $\rho = min\{34, 19, 34\} = 19$. Ко всем клеткам со знаком «+» или «-» разносим груз. Маршрут x_{42} закрывается. Стоимость нового плана $F = 277$ грн.

Опорность плана при записи условий транспортной задачи в виде таблицы заключается в его ацикличности, т.е. в таблице нельзя построить замкнутый цикл, все вершины которого лежит в занятых клетках. Всякий план транспортной задачи, содержащий более $m + n - 1$ занятых клеток, не является опорным, поскольку ему соответствует линейно зависимая система векторов.

Постав- щики	Потребители			
	B_1	B_2	B_3	B_4
A_1	4 0	3 16	2 30	7 0
A_2	1 15	1 19	6 0	0 0
A_3	3 25	5 0	9 0	4 15
A_4	0 0	0 0	0 0	0 30

$$F = 3 \cdot 16 + 2 \cdot 30 + 1 \cdot 15 + 19 \cdot 1 + 3 \cdot 25 + 15 \cdot 4 = 277 \text{ грн.}$$

	$v_1 = 3$	$v_2 = 3$	$v_3 = 2$	$v_4 = 4$

$u_1 = 0$	3 / 4	3	2	4 / 7
$u_2 =$-2	1	1	0 /	2 / 4
$u_3 =$0	3	3 / 5	2 / 9	4
$u_4 =$-4	-1 / 0	-1 / 0	-2 / 0	0

Вывод: план оптимальный, поскольку все $\Delta_{ij} = (u_i + v_j) - c_{ij} < 0$.

Экономическая оценка $\Delta_{ij} = c_{ij} - (u_i + v_j) < 0$ показывает, на сколько денежных единиц уменьшаться транспортные издержки от загрузки данной клетки единицей груза. Эффективность плана от загрузки потенциальной клетки грузом ρ единиц составляет $\Delta F = \Delta_{ij} \cdot \rho$ денежных единиц.

8.3. Вырождение плана транспортной задачи

При решении транспортной задачи может оказаться, что число занятых клеток меньше, чем $r = m + n - 1$. В этом случае задача имеет вырожденное решение. Для возможного его исключения целесообразно поменять местами поставщиков и потребителей или ввести в свободную клетку с наименьшей стоимостью нулевую поставку. Ноль помещают в такую клетку, чтобы в каждой строке и каждом столбце было не менее одной занятой клетки.

Пример. Фирма осуществляет поставку бутылок на три завода, занимающиеся производством прохладительных напитков. Она имеет три склада, причем на складе 1 находится 6000 бутылок, на складе 2 – 3000 бутылок и на складе 3 – 4000 бутылок. Первому заводу требуется 4000 бутылок, второму заводу – 5000 бутылок, третьему – 1000 бутылок, четвертому – 3000. Матрицей $\begin{pmatrix} 6 & 4 & 9 & 8 \\ 5 & 3 & 2 & 8 \\ 2 & 3 & 6 & 8 \end{pmatrix}$ задана стоимость перевозки одной бутылки от каждого склада к каждому заводу. Необходимо так организовать доставку бутылок на заводы, чтобы стоимость перевозки была минимальной.

Решение. Исходное опорное решение получим по методу минимальной стоимости.

Склады	Заводы				
	1	**2**	3	4	
1	6 0	4 3000	9 0	8 3000	6000

2	5 0	3 2000	2 1000	8 0	3000
3	2 4000	3 0	6 0	8 0	4000
	4000	5000	1000	3000	

Число заполненных клеток 5, $r = m + n - 1 = 6$. Следовательно задача является вырожденной.

Для исключения вырожденности необходимо в какую-то клетку ввести нулевую поставку. Такая клетка становится условно занятой, ее целесообразно определить при вычислении потенциалов занятых клеток, она должна иметь наименьшую стоимость по сравнению с другими клетками, которые могут быть условно занятыми. Для нахождения потенциалов поместим нулевую поставку в клетку (3, 2).

	$v_1 = 3$	$v_2 = 4$	$v_3 = 3$	$v_4 = 8$
$u_1 = 0$	3 6	**4**	3 9	8
$u_2 = -1$	**2** **5**	3	2	7 8
$u_3 = -1$	2	3	2 6	7 8

План оптимальный. Стоимость транспортных расходов будет минимальной и составит $F = 4 \cdot 3000 + 8 \cdot 3000 + 3 \cdot 2000 + 2 \cdot 1000 + 2 \cdot 4000 = 52000$ ден.ед.

Вопросы для самопроверки:

1) Описать экономико-математическую модель транспортной задачи.

2) Записать транспортную задачу в матричной форме и назвать ее свойства.

3) Назвать методы решения транспортной задачи.

4) Какой критерий оптимальности для транспортной задачи в матричной форме?

5) Дать экономическую интерпретацию метода потенциалов решения транспортной задачи.

6) Как осуществить переход к другому базисному плану?

7) Как решить транспортную задачу в случае вырождения?

Тема 9. Задачи экономического содержания, сводящиеся к транспортной задаче

9.1. Приложение транспортных моделей к решению некоторых экономических задач.

9.2. Выбор оптимального варианта использования производственного оборудования.

9.1. Приложение транспортных моделей к решению некоторых экономических задач.

Алгоритмы и методы решения транспортной задачи могут быть использованы при решении некоторых экономических задач, не имеющих ничего общего с транспортировкой груза. В этом случае величины стоимости c_{ij} имеют различный смысл в зависимости от конкретной экономической задачи. К таким относятся следующие задачи:

- оптимальное закрепление за станками операций по обработке деталей. В них c_{ij} является таким экономическим показателем, как производительность. Задача позволяет определить, сколько времени и на какой операции нужно использовать каждый из станков, чтобы обработать максимальное количество деталей. Так как транспортная задача требует нахождения минимума, то значения c_{ij} берутся с отрицательным знаком;

- оптимальные назначения или проблема выбора. Имеется m механизмов, которые могут выполнять m различных работ с производительностью c_{ij}. Задача позволяет определить, какой механизм и на какую работу надо назначить, чтобы добиться максимальной производительности;

- задача о сокращении производства с учетом суммарных расходов на изготовление и транспортировку продукции;

- увеличение производительности автомобильного транспорта за счет минимизации порожнего пробега. Уменьшение порожнего пробега сократит количество автомобилей для перевозок, увеличив их производительность;

- решение задач с помощью метода запрещения перевозок. Используется в том случае, если груз от некоторого поставщика по каким-то причинам не может быть направлен одному из потребителей. Данное ограничение можно учесть, присвоив соответствующей клетке достаточно большое значение стоимости, тем самым в этой клетку не будут производиться перевозки.

9.2. Выбор оптимального варианта использования производственного оборудования

Пример. На предприятии имеются три группы станков, каждая из которых может выполнять пять операций по обработке деталей (операции могут выполняться в любом порядке). Максимальное время работы каждой группы станков соответственно равно 100, 250, 180 часов. Каждая операция должна выполняться соответственно 100, 120, 70, 110, 130 ч.

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

Производительность каждой группы станков на каждую операцию задана матрицей

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

Решение.

Поскольку в задаче требуется найти максимум, а в классической транспортной задаче находится минимум, тарифы следует умножить на (-1).

Станки	Операции					Время
	B_1	B_2	B_3	B_4	B_5	
A_1	-3 40	-5 0	-11 0	-10 0 +	-5 60 -	100
A_2	-5 60	-10 120	-15 70	-3 0	-2 0	250
A_3	-4 0	-8 0	-6 0	-12 110-	-10 70 +	180
Время	100	120	70	110	130	530

$-F = 3\cdot 40 + 5\cdot 60 + 5\cdot 60 + 10\cdot 120 + 15\cdot 70 + 12\cdot 110 + 10\cdot 70 = 5030$ шт деталей

	$v_1 = -3$	$v_2 = -8$	$v_3 = -13$	$v_4 = -7$	$v_5 = -5$
$u_1 = 0$	-3	-8 -5	-13 -11	-7 -10	-5
$u_2 = -2$	-5	-10	-15	-9 -3	-7 -2
$u_3 = -5$	-8 -4	-13 -8	-18 -6	-12	-10

Поскольку $\Delta_{14} = 3 > 0$, то требуется перераспределить время.

Станки	Операции					Время
	B_1	B_2	B_3	B_4	B_5	
A_1	-3 40	-5 0	-11 0	-10 60	-5 0	100
A_2	-5 60	-10 120	-15 70	-3 0	-2 0	250
A_3	-4 0	-8 0	-6 0	-12 50	-10 130	180
Время	100	120	70	110	130	530

$-F = 3 \cdot 40 + 5 \cdot 60 + 10 \cdot 60 + 10 \cdot 120 + 15 \cdot 70 + 12 \cdot 50 + 10 \cdot 130 = 5170$ шт деталей

Проверим новый план на оптимальность.

	$v_1 =$-3	$v_2 =$-8	$v_3 =$-13	$v_4 =$ -10	$v_5 =$ -8
$u_1 = 0$	-3	-8 -5	-13 -11	-10	-8 -5
$u_2 =$-2	-5	-10	-15	-12 -3	-10 -2
$u_3 =$-2	-5 -4	-10 -8	-15 -6	-12	-10

План оптимальный. Таким образом, на первой группе станков целесообразно выполнять операции 1 и 4 продолжительностью 40 и 60 ч соответственно, а на второй группе – операции 1, 2 и 3 продолжительностью 60, 120 и 70 ч соответственно, на третьей группе – операции 4 и 5 продолжительностью 50 и 130 ч соответственно. При этом максимальное число обработанных деталей составит 5170 штук.

Вопросы для самопроверки:

1) Привести примеры экономических задач, для которых целесообразно использовать алгоритмы и методы решения транспортной задачи.

2) В чем особенность решения задачи выбора оптимального варианта использования производственного оборудования.

Тема 10. Задачи дробно-линейного программирования. Основные методы их решения и анализа

10.1. Экономическая и математическая постановка задачи дробно-линейного программирования.

10.2. Геометрическая интерпретация задачи дробно-линейного программирования.

10.3. Решение дробно-линейной задачи сведением к задаче линейного программирования.

10.1. Экономическая и математическая постановка задачи дробно-линейного программирования.

При решении экономических задач в качестве критерия оптимальности берут уровень рентабельности, производительности труда и т.д. Эти показатели математически выражаются дробно-линейными функциями. В этом случае общая экономико-математическая модель следующая:

обозначим c_j- прибыль от реализации единицы j-го вида продукции, тогда общая прибыль $\sum_{j=1}^{n} c_j x_j$; d_j- затраты на производство единицы j-го вида продукции, тогда $\sum_{j=1}^{n} d_j x_j$ - общие затраты на производство. В случае максимизации уровня рентабельности производства целевая функция имеет вид:

$$Z = \frac{\sum_{j=1}^{n} c_j x_j}{\sum_{j=1}^{n} d_j x_j} \to \max \quad (9.1\ )$$

при условии выполнения ограничений относительно использования ресурсов:

$$\sum_{j=1}^{n} a_{ij} x_j \ (\le, =, \ge)\ b_i$$
$$(i = \overline{1,m}), \quad x_j \ge 0 \quad (j = \overline{1,n}).$$

Знаменатель целевой функции в области допустимых решений системы ограничений не равен нулю.

Поскольку задача (9.1) отличается от обычной задачи линейного программирования лишь целевой функцией, поэтому для ее решения используются модифицированные известные методы решения задач линейного программирования.

10.2. Геометрическая интерпретация задачи дробно-линейного программирования.

Пусть дана задача дробно-линейного программирования для случая двух переменных:

$$Z = \frac{c_1 x_1 + c_2 x_2}{d_1 x_1 + d_2 x_2} \to \max$$

$$\begin{cases} a_{11} x_1 + a_{12} x_2 \le b_1, \\ a_{21} x_1 + a_{22} x_2 \le b_2, \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ a_{m1} x_1 + a_{m2} x_2 \le b_m, \\ x_1 \ge 0, \quad x_2 \ge 0 \end{cases}$$

Как и в линейном программировании, строим область допустимых решений согласно системе ограничений. Для построения целевой функции ее выражение следует упростить:

$$\frac{c_1 x_1 + c_2 x_2}{d_1 x_1 + d_2 x_2} = Z,$$

$$(c_1 - Z d_1) x_1 + (c_2 - Z d_2) x_2 = 0,$$

$$x_2 = -\frac{(c_1 - Z d_1)}{(c_2 - Z d_2)} x_1,$$

$$x_2 = k\, x_1,$$

$$k = \frac{(Z d_1 - c_1)}{(c_2 - Z d_2)}$$

Уравнение $x_2 = k\, x_1$ представляет прямую с угловым коэффициентом, проходящую через начало координат. Если Z принимает постоянное значение ($Z = const$), то некоторое постоянное значение получит и коэффициент k, а прямая займет на плоскости в прямоугольной системе координат определенное положение (рис. 1). При другом постоянном значении Z прямая $x_2 = k\, x_1$ займет иное положение. Оно как бы повернется вокруг начала координат.

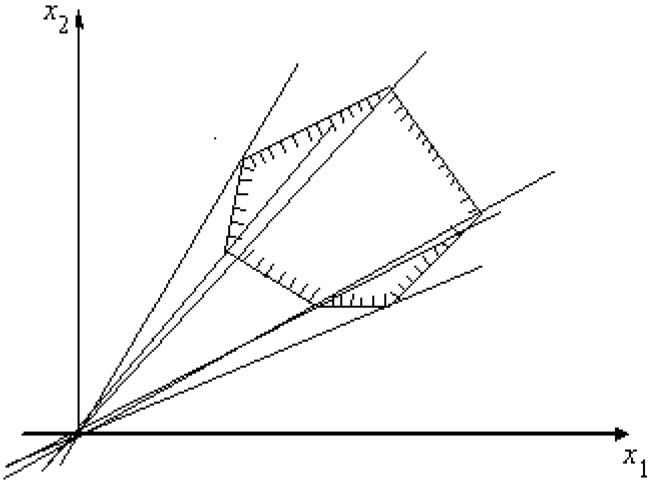


Рис.1.

Выясним знак производной углового коэффициента k по отношению к Z:

$$k'(Z) = \frac{(c_2 d_1 - c_1 d_2)}{(c_2 - Z d_2)^2}.$$

Отсюда видно, что знаменатель дроби всегда положителен, а числитель не зависит от целевой функции Z. Следовательно, производная k' имеет постоянный знак.

Правило 1. Если $k'(Z) = \frac{(c_2 d_1 - c_1 d_2)}{(c_2 - Z d_2)^2} > 0$, $(c_2 d_1 - c_1 d_2) > 0$, для отыскания точки максимума необходимо поворачивать прямую, которая описывает целевую функцию около начала координат в направлении против часовой стрелки.

Правило 2. Если $k'(Z) = \frac{(c_2 d_1 - c_1 d_2)}{(c_2 - Z d_2)^2} < 0$, $(c_2 d_1 - c_1 d_2) < 0$ - за часовой стрелкой.

Пример. На плодоконсервном заводе из трех видов фруктов (яблоки, груши, сливы) изготавливают компот двух видов. Количество фруктов, необходимое для приготовления 1 л компота, запас фруктов и затраты даны в таблице. Известно также, что яблок можно расходовать не более 500 кг, груш и слив – не менее 400 и 300 кг соответственно. Требуется составить план производства компота двух видов и получить максимальную и минимальную себестоимость 1 л компота.

Фрукты	Запасы, кг	Расход (кг) на компот вида	
		I	II
Яблоки	500	0,3	0,2
Груши	400	0,2	0,3
Сливы	300	0,3	0,2
Затраты на 1 л компота, ден. ед.		0,5	0,6

Решение. Обозначим x_1 - количество литров компота I вида, x_2 - количество литров компота II вида. Тогда система ограничений по расходу фруктов примет вид:

$$\begin{cases} 0,3x_1 + 0,2x_2 \le 500 \\ 0,2x_1 + 0,3x_2 \ge 400 \\ 0,3x_1 + 0,2x_2 \ge 300 \end{cases}$$
$$x_1 \ge 0,\ x_2 \ge 0$$

Затраты на производство компота: $z_1 = 0,5x_1 + 0,6x_2$, общий объем производства компота представляется функцией $z_2 = x_1 + x_2$, отношение общих затрат к общему объему производства дает себестоимость 1 л компота:

$$Z = \frac{0,5x_1 + 0,6x_2}{x_1 + x_2} \to \min,$$

$$\begin{cases} 0,3x_1 + 0,2x_2 \le 500 \\ 0,2x_1 + 0,3x_2 \ge 400 \\ 0,3x_1 + 0,2x_2 \ge 300 \end{cases}$$
$$x_1 \ge 0,\ x_2 \ge 0$$

Строим область допустимых решений (ОДР). Чтобы определить, в какой точке ОДР целевая функция будет иметь экстремальное значение, выразим x_2:

$$x_2 = \frac{0,5 - Z}{Z - 0,6} x_1.$$

Угловой коэффициент

$$k = \frac{0,5 - Z}{Z - 0,6},$$

Его производная

$$k' = \frac{0,1}{(Z - 0,6)^2}.$$

Поскольку производная при любом значении Z положительна, функция $k = \dfrac{0,5 - Z}{Z - 0,6}$ возрастающая; с увеличением Z угловой коэффициент увеличивается. Это соответствует вращению прямой против хода часовой стрелки.

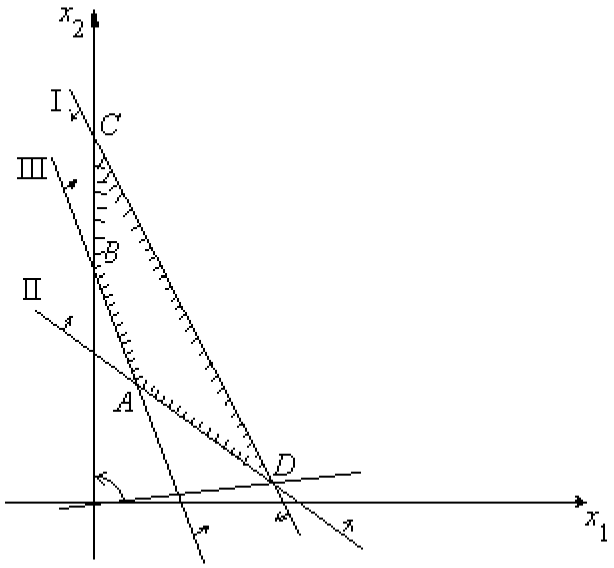


Рис. 2.

Таким образом, наибольшее значение целевая функция будет достигать в вершинах B и C области допустимых решений, т.е. максимальная себестоимость определяется во всех точках ребра BC. Координаты точек B и C равны соответственно (0; 1500) и (0; 2500); целевая функция в этих точках:

$$Z_{\max} = Z(B) = Z(C) = 0,6,$$

Т.е. максимальная себестоимость 1 л компота составляет 60 ден. ед.

Минимальная себестоимость достигается в точке D области допустимых решений; целевая функция в точке D

$$Z_{\min} = Z(D) \approx 0,52.$$

Следовательно, минимальная себестоимость 1 л компота 52 ден. ед, максимальная – 60 ден. ед.

На рис. 2 видно, что себестоимость 1 л компота остается постоянной, равной 60 ден. ед при производстве компота II вида от 1500 до 2500 л., т.е. $1500 \le x_2 \le 2500$.

Замечание. Решение задач графическим методом может быть проще. Достаточно найти координаты крайних точек области допустимых решений, а затем вычислить значения целевой функции в этих точках и выбрать из них наибольшее и наименьшее.

10.3. Решение дробно-линейной задачи сведением к задаче линейного программирования.

Задачу дробно-линейного программирования можно свести к задаче линейного программирования и решить симплексным методом.

Обозначим $y_0 = \dfrac{1}{\sum\limits_{j=1}^{n} d_j x_j}$ при условии $\sum\limits_{j=1}^{n} d_j x_j \neq 0$ и введем новые переменные $y_j = y_0 x_j$. Тогда задача примет вид $F = \sum\limits_{j=1}^{n} c_j y_j \to \max(\min)$ при ограничениях

$$
\begin{aligned}
&\sum_{j=1}^{n} a_{ij} y_j - b_i y_0 = 0 \\
&\sum_{j=1}^{n} d_j y_j = 1 \\
&(i = \overline{1,m}), \quad y_j \geq 0 \quad (j = \overline{1,n}), \quad y_0 > 0.
\end{aligned}
$$

После нахождения оптимального решения полученной задачи, используя вышеуказанные соотношения, находят оптимальное решение исходной задачи дробно-линейного программирования.

Пример. Дана задача дробно-линейного программирования:

$F = \dfrac{2x_1 - x_2}{x_1 + 2x_2 + 1} \to \max$ при ограничениях:

$$
\begin{cases}
x_1 - 2x_2 + x_3 = 2 \\
2x_1 + x_2 + x_4 = 6. \\
x_j \geq 0, \ j = \overline{1,4}
\end{cases}
$$

Решение. Обозначим $x_1 + 2x_2 + 1 = \dfrac{1}{y_0}$, $y_0 > 0$, тогда $F = 2x_1 y_0 - x_2 y_0$.

Обозначим: $x_1 y_0 = y_1$, $x_2 y_0 = y_2$, $x_3 y_0 = y_3$, $x_4 y_0 = y_4$.

Преобразуем систему ограничений, умножив обе части всех ограничений на y_0, и перейдем к переменным y_0, y_1, y_2, y_3, y_4. Задача примет вид $F = 2y_1 - y_2 \to \max$ при ограничениях

$$\begin{cases} -2y_0 + y_1 - 2y_2 + y_3 = 0 \\ -6y_0 + 2y_1 + y_2 + y_4 = 0 \\ y_0 + y_1 + 2y_2 = 1 \\ y_j \ge 0,\ j = \overline{1,4},\ y_0 > 0. \end{cases}$$

№	Базис	$c_{баз}$	c_j	0	2	-1	0	0	Примечания
			A_0	y_0	A_1	A_2	A_3	A_4	
1	A_3	0	0	-2	1	-2	1	0	
2	A_4	0	0	-6	2	1	0	1	
3			1	1	1	2	0	0	
4	A_3	0	2	0	3	2	1	0	
5	A_4	0	6	0	8	13	0	1	
6	y_0	0	1	1	1	2	0	0	
7	Δ_1		0	0	-2	1	0	0	
8	A_1	2	2/3	0	1	2/3	1/3	0	
9	A_4	0	2/3	0	0	2/3	-8/3	1	
10	y_0	0	1/3	1	0	4/3	-1/3	0	
11	Δ_2		4/3	0	0	7/3	2/3	0	

Имеем $Y^* = (1/3, 2/3, 0, 0, 2/3)$,

Тогда $x_1 = \dfrac{y_1}{y_0} = 2, \quad x_2 = \dfrac{y_2}{y_0} = 0, \quad x_3 = \dfrac{y_3}{y_0} = 0, \quad x_4 = \dfrac{y_4}{y_0} = 2,$

$X^* = (2, 0, 0, 2), \quad F_{\max} = 4/3.$

Вопросы для самопроверки:

1) Привести экономическую и математическую постановку задачи дробно-линейного программирования.

2) Какие методы применяются для решения задачи дробно-линейного программирования.

3) Какая геометрическая интерпретация задачи дробно-линейного программирования?

4) Как решение дробно-линейной задачи сводится к задаче линейного программирования?

Тема 11. Целочисленные задачи линейного программирования. Основные методы их решения и анализа

11.1. Экономическая постановка задачи целочисленного программирования и ее математическая модель.

11.2. Графический метод решения задач.

11.3. Основные методы решения целочисленных задач. Геометрическая интерпретация решений целочисленной задачи на плоскости. Метод Гомори.

11.4. Метод вервей и границ

11.1. Экономическая постановка задачи целочисленного программирования и ее математическая модель.

Значительная часть задач коммерческой деятельности требует целочисленного решения. К ним относятся задачи, у которых переменные величины означают количество единиц неделимой продукции, например, распределения товаров между предприятиями, раскрой материалов, число станков при загрузке оборудования, распределение транспортных средств по рейсам, продажа автомобилей, распределение самолетов по авиалиниям, количество персональных компьютеров в управляющем комплексе и т.д. Линейные задачи, решения которых должно быть получено в целых числах, называют задачами целочисленного программирования.

В общем виде математическая модель целочисленного программирования имеет вид:

$$Z = \sum_{j=1}^{n} c_j x_j \to \max (\min) \quad (1)$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad (2)$$

$$i = \overline{1,m}, \; j = \overline{1,n}$$

$$x_j \geq 0, \quad x_j \text{ – целые.} \quad (3)$$

11.2. Графический метод решения задач.

При наличии в задаче линейного программирования двух переменных, а в системе ограничений – неравенств, она может быть решена графическим методом.

В системе координат X_1OX_2 находят область допустимых решений, строят вектор $\overline{C}$ и линию уровня. Перемещая линию уровня по направлению $\overline{C}$ для задач на максимум, находим наиболее удаленную от начала координат точку и ее координаты.

В том случае, когда координаты этой точки целочисленные, в области допустимых решений строят целочисленную решетку и находят на ней такие целые числа, которые удовлетворяют системе ограничений и при которых значение целевой функции наиболее близко к экстремальному нецелочисленному решению. Координаты такой вершины и являются целочисленным решением.

Аналогично решается задача на минимум.

Пример. Для улучшения финансового положения фирма приняла решение об увеличении выпуска конкурентоспособной продукции, для чего принято решение об установке в одном из цехов дополнительного оборудования, занимающего 19/3 м^2 площади. На приобретение дополнительного оборудования фирма выделила 10 усл. Ед., при этом она может купить оборудование двух видов. Приобретение 1-го комплекта оборудования 1-го вида стоит 1,0 усл.ед., 2-го вида – 3 усл.ед. Приобретение одного комплекта оборудования 1-го вида позволяет увеличить выпуск продукции в смену на 2 шт., а одного комплекта оборудования 2-го вида – на 4 шт. Зная, что установки одного комплекта оборудования 1-го вида требуется 2 м^2 площади, а для оборудования 2-го вида – 1 м^2 площади, определить такой набор дополнительного оборудования, который дает возможность максимально увеличить выпуск продукции.

Решение. Составим математическую модель задачи.

$$Z = 2x_1 + 4x_2 \to \max$$

$$\begin{cases} 2x_1 + x_2 \le 19/3, \\ x_1 + 3x_2 \le 10, \\ x_1, x_2 - \text{целые.} \end{cases}$$

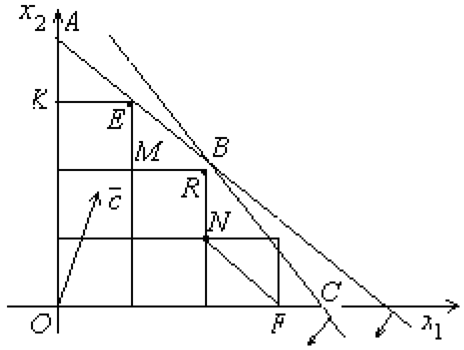


Рис.2.

Без условия целочисленности, четырехугольник $OABC$ – область допустимых решений. Оптимальное решение в точке $B\left(\frac{9}{5}, \frac{41}{15}\right)$, $Z_{\max} = \frac{218}{15}$ усл. ед.

Условию целочисленности удовлетворяют 12 точек. Заменяем многоугольник $OABC$ на $OKEMRNF$, содержащем все допустимые точки с целочисленными координатами. По градиенту оптимум в точке $E(1, 3)$, $Z_{\max}(X_{\text{цел}}) = 2 \cdot 1 + 4 \cdot 3 = 14$ усл. ед.

Вывод. Фирме следует приобрести один комплект оборудования 1-го вида и три комплекта оборудования 2-го вида, что обеспечит ей при имеющихся ограничениях на производственные площади и денежные средства максимальное увеличение выпуска продукции, равное 14 усл. ед в смену.

11.3. Основные методы решения целочисленных задач. Геометрическая интерпретация решений целочисленной задачи на плоскости. Метод Гомори.

Методы целочисленной оптимизации можно разделить на три основные группы: а) метод отсечения; б) комбинированные методы; в) приближенные методы.

Рассмотрим один из методов отсечения – метод Гомори.

Алгоритм Гомори:

1) решаем задачу без учета условий целочисленности;

2) полученное оптимальное решение (если оно существует) проверяется на целочисленность. Если условие целочисленности выполняется по всем переменным, то оптимальное решение найдено. Если это условие не выполняется, то переходим к следующему этапу или делаем вывод, что задача оказалась неразрешимой (случай: базисная переменная дробная, а все коэффициенты в строке целые числа);

3) строится дополнительное ограничение, отсекающее часть области, в которой содержится оптимальное решение задачи (1)-(2) и не содержится ни одного допустимого решения задачи (1)-(3);
4) последний этап предусматривает возвращение к задаче линейного программирования с отброшенным условием целочисленности, но с расширенной системой ограничений, в которую включено дополнительное ограничение, полученное на 3-ем шаге. К расширенной системе ограничений вновь применяется симплексная процедура. Если найденное таким образом решение будет опять нецелочисленным, то формируется новое дополнительное ограничение и процесс вычислений повторяется.

Алгоритм Гомори позволяет за конечное число шагов прийти к оптимальному целочисленному решению, если оно существует. Геометрическая интерпретация решений целочисленной задачи методом Гомори на плоскости представлена рис. 1.

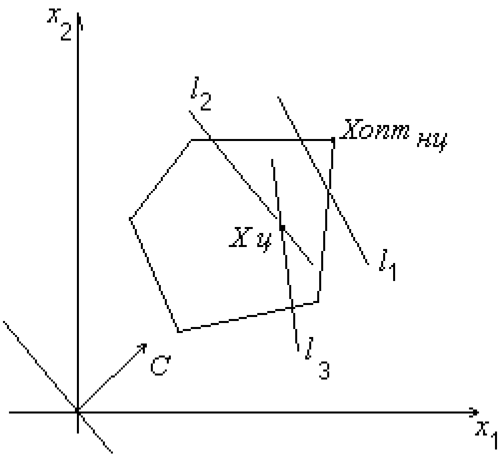


Рис. 1.

Где $X_{опт\ нц}$ - оптимальное решение без учета целочисленности; три прямые l_1, l_2, l_3 соответствуют трем дополнительным линейным ограничениям.

Требуется составить правильное дополнительное ограничение, которое отсекает часть области допустимых решений. Правило отсечения должно удовлетворять следующим усливиям 1) быть линейным; 2) отсекает найденное оптимальное нецелочисленное решение задачи; 3) не отсекает ни одной из целочисленных точек задачи (1)-(3).

Процедура формирования правильного отсечения:

После каждой итерации система ограничений имеет вид:

$$x_i = \beta_i - \sum_{x_j \in (НП)} \alpha_{ij} x_j$$
$$x_i \in (БП) - \textit{множество базисных переменных}.$$

Если выполняется условие оптимальности задачи, то находим оптимальное решение. Если все компоненты оптимального плана целочисленны, то задача решена.

Пусть некоторые β_0- нецелые, пусть компонента i_0-нецелая. Рассмотрим равенство, для которого выполняется условие оптимальности:

$$x_{i0} = \beta_{i0} - \sum_{x_j \in (НП)} \alpha_{i0} x_j .$$

Число представляется как: $\alpha = [\alpha] + \{\alpha\},\ 0 \le \{\alpha\} < 1$, тогда

$\beta_{i0} = [\beta_{i0}] + \{\beta_{i0}\},\ \alpha_{i0j} = [\alpha_{i0j}] + \{\alpha_{i0j}\},\ x_j \in (НП)$.

Пусть β_{i0} нецелое, то $\{\beta_{i0}\} > 0$ и $\{\alpha_{i0j}\} \ge 0$, тогда

$$x_i = \left([\beta_{i0}] - \sum_{x_j \in (НП)} [\alpha_{i0j}] x_j \right) + \left(\{\beta_{i0}\} - \sum_{x_j \in (НП)} \{\alpha_{i0j}\} x_j \right). \qquad (4)$$

Поскольку первое слагаемое равенства (4) есть целое число, то, для того чтобы x_{i0} было целым, необходимо, чтобы второе слагаемое также было целым, т.е. величина $L_{i0} = \{\beta_{i0}\} - \sum_{x_j \in (НП)} \{\alpha_{i0j}\} x_j$ должна быть целым числом. Покажем, что $L_{i0} \le 0$ - целое число. Величина $\sum_{x_j \in (НП)} \{\alpha_{i0j}\} x_j$ не может быть отрицательной; из условия $0 \le \{\beta_{i0}\} < 1$. И учитывая, что L_{i0}-целое, то если $L_{i0} > 0$, следует, что $\{\beta_{i0}\} > 1$- это противоречит определению дробной части числа.

Таким образом, доказано, что любое допустимое решение задачи (1)-(3) должно удовлетворять неравенству $\{\beta_i\} - \sum_{x_j \in (НП)} \{\alpha_{ij}\} x_j \le 0$.

Теорема. Неравенство $\{\beta_i\} - \sum_{x_j \in (НП)} \{\alpha_{ij}\} x_j \le 0$ определяет правильное отсечение Гомори, т.е.: 1) является линейным; 2) отсекает найденное оптимальное нецелочисленное решение задачи; 3) не отсекает ни одного целочисленного плана задачи.

Продолжение примера. Решить задачу целочисленного программирования методом Гомори.

$$Z = 2x_1 + 4x_2 \to \max$$

$$\begin{cases} 2x_1 + x_2 \le 19/3, \\ x_1 + 3x_2 \le 10, \\ x_1, x_2 - \text{целые}. \end{cases} \to \begin{cases} 2x_1 + x_2 + x_3 = 19/3, \\ x_1 + 3x_2 + x_4 = 10, \\ x_1, x_2 - \text{целые}. \end{cases}$$

№	Базис	$c_{баз}$	c_j	2	4	0	0	0	Примечания
			A_0	A_1	A_2	A_3	A_4	A_5	
1	A_3	0	19/3	2	1	1	0		
2	A_4	0	10	1	3	0	1		
		Δ_1	0	-2	-4	0	0		
3	A_3	0	3	5/3	0	1	-1/3		
4	A_2	4	10/3	1/3	1	0	1/3		
	Δ_2		40/3	-2/3	0	0	4/3		
5	A_1	2	9/5	1	0	3/5	-1/5	0	
6	A_2	4	41/15	0	1	-1/5	2/5	0	
		Δ_3	218/15	0	0	2/5	6/5		
7	A_5	0	-4/5	0	0	-3/5	-4/5	1	
8	A_1	2	1	1	0	0	-1	1	
9	A_2	4	3	0	1	0	2/3	-1/3	
10	A_3	0	4/3	0	0	1	4/3	-5/3	
11		Δ_3	14	0	0	0	2/3	2/3	

Нецелочисленное оптимальное решение:

$$X^*\left(\frac{9}{5}, \frac{41}{15}\right), \; Z_{\max} = \frac{218}{15} \text{ усл. ед.}$$

Формируем дополнительное ограничение:

найдем дробные части чисел 9/5 и 41/15

$\{a\} = a - [a]$;

$\left\{\frac{9}{5}\right\} = \frac{9}{5} - 1 = \frac{4}{5}$, $\left\{\frac{41}{15}\right\} = \frac{41}{15} - 2 = \frac{11}{15}$, $\frac{4}{5} = \frac{12}{15} > \frac{11}{15}$ следовательно, составляем дополнительное ограничение для 5-й строки. Найдем дробные части $\left\{\frac{3}{5}\right\} = \frac{3}{5} - 0 = \frac{3}{5}$, $\left\{-\frac{1}{5}\right\} = -\frac{1}{5} - (-1) = -\frac{1}{5} + 1 = \frac{4}{5}$, дополнительное ограничение будет иметь вид:

$\frac{3}{5}x_3+\frac{4}{5}x_4 \geq \frac{4}{5}$ или $\frac{3}{5}x_3+\frac{4}{5}x_4 - x_5 = \frac{4}{5}$ и добавляем в симплекс-таблицу (строка 7).

<u>Целочисленное оптимальное решение:</u>

$X_{цел}^{*}(\ ,3)$, $Z_{\max}=14$ усл. ед.

11.4. Метод вервей и границ

Основы этого метода целесообразно рассмотреть на численном примере.

Пример. Найти

$$Z=5x_1+4x_2 \rightarrow max$$

$$\begin{cases} x_1+x_2 \leq 5, \\ 10x_1+6x_2 \leq 45, \\ x_1, x_2 \geq 0 \text{ и целые.} \end{cases}$$

Решение. Начальную задачу линейного программирования (ЛП0) (т.е. без учета целочисленности) решим графически (рис. 2). Ее оптимальным решением будет $x_1=3{,}75;\ x_2=1{,}25;\ z=23{,}75.$

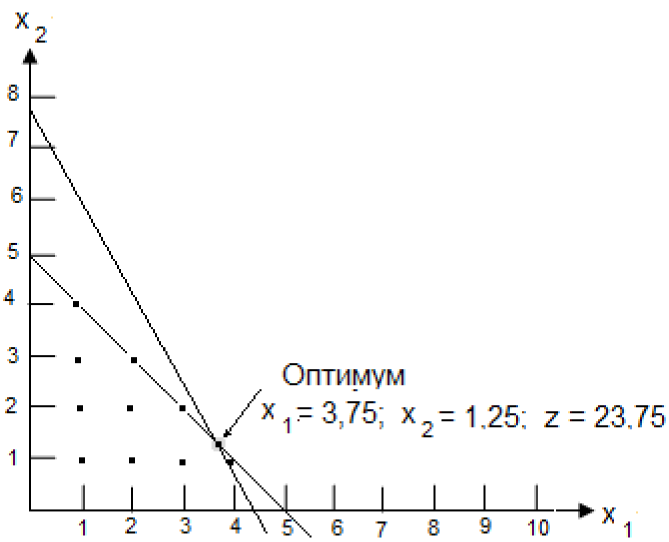


Рис.2

Поскольку оптимальное решение не целочисленное, применим метод ветвей и границ, который изменяет пространство решений задачи линейного программирования. Сначала выбирается одна из переменных, которая должна быть целочисленной и значение которой в оптимальном решении нецелочисленное. Выберем $x_1=3{,}75$. Область $3<x_1<4$ пространства допустимых решений задачи ЛП0 не содержит целочисленных значений переменной x_1 и может быть исключена из рассмотрения. Это эквивалентно замене исходной задачи ЗЛ0 двумя новыми задачами линейного программирования ЛП1 и ЛП2, которые определяются так:

1) пространство допустимых решений ЛП1 - пространство допустимых решений ЛП0 + ($x_1 \le 3$),

2) пространство допустимых решений ЛП2 - пространство допустимых решений ЛП0 + ($x_1 \ge 4$).

На рис. 3 изображены пространства допустимых решений задач ЛП1 и ЛП2.

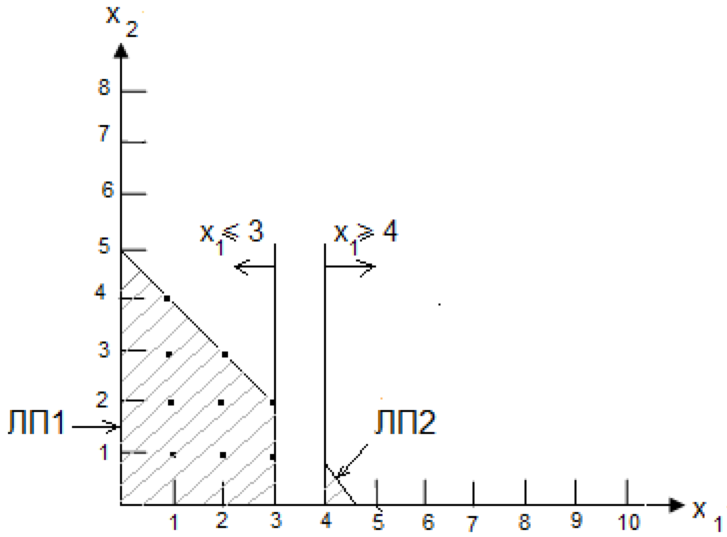


Рис. 3

Далее будем задачу целочисленного программирования путем решения последовательности непрерывных задач линейного программирования.

Новые ограничения $x_1 \le 3$ и $x_1 \ge 4$ взаимоисключаемы, поскольку ЛП1 и ЛП2 необходимо рассматривать как независимые задачи линейного программирования. Дихотомизация задач ЛП – основа концепции ветвления в методе ветвей и границ. В этом случае x_1 называется переменной ветвления.

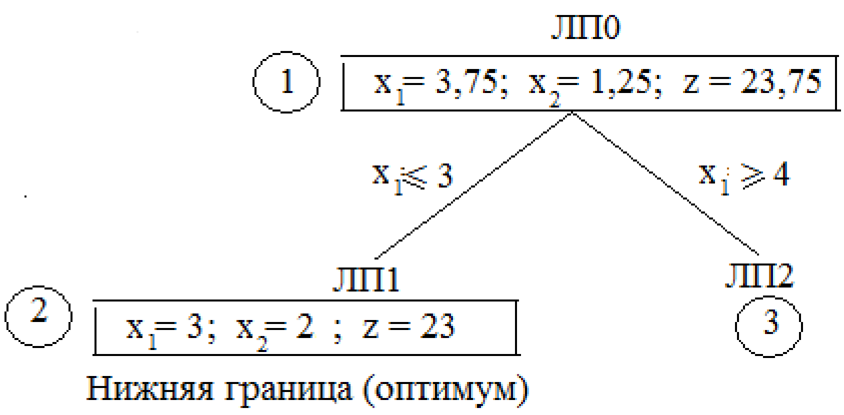


Рис. 4

Оптимальное решение задачи ЦЛП находится в пространстве допустимых решений либо задачи ЛП1, либо ЛП2. Следовательно, обе подзадачи должны быть решены. Выбираем сначала задачу ЛП1 (выбор произволен), имеющую дополнительное ограничение $x_1 \le 3$, т.е.

$$Z = 5x_1 + 4x_2 \to max$$
$$\begin{cases} x_1 + x_2 \le 5, \\ 10x_1 + 6x_2 \le 45, \\ x_1 \le 3, \\ x_1, x_2 \ge 0 \text{ и целые.} \end{cases}$$

Оптимальным решением задачи ЛП1 является $x_1 = 3, x_2 = 2, z = 23$. Оптимальное решение задачи ЛП1 удовлетворяет требованию целочисленности переменных, в этом случае говорят, что задача ЛП1 прозондирована и можем сказать, что значение $z = 23$ является нижней границей оптимального (максимального) значения целевой функции исходной задачи ЦЛП. При значении нижней границы $z = 23$ исследуем задачу ЛП2. Так как в задаче ЛП0 оптимальное значение целевой функции равно 23,75 и все ее коэффициенты являются целыми числами, то невозможно получить целочисленное решение задачи ЛП2 (пространство решений которой более узко, чем в задаче ЛП0), которое будет лучше имеющегося. В результате мы отбрасываем подзадачу ЛП2 и считаем ее прозондированной.

Рассмотрим два момента: 1) при выборе подзадачи можно для зондирования решить сначала задачу ЛП2 вместо ЛП1; 2) в задаче ЛП0 сначала выбрать переменную x_2 в качестве переменной ветвления вместо x_1.

Ситуация, когда первой решается задача ЛП2, иллюстрируется схемой вычислений (рис. 5). Поскольку значение переменной $x_2 = 0{,}83$ не является целым числом, задача ЛП2 исследуется дальше. Рассматриваем задачи ЛП3 и ЛП4, используя ветви $x_2 \le 0$ и $x_2 \ge 0$ соответственно. Это означает, что

1) пространство решений ЛП3 = пространство решений ЛП2 + ($x_2 \le 0$) = пространство решений ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$).

2) пространство решений ЛП4 = пространство решений ЛП2 + ($x_2 \ge 1$) = пространство решений ЛП0 + ($x_1 \ge 4$) + ($x_2 \ge 1$).

Есть еще три нерассмотренные задачи, которые должны быть решены – ЛП1, ЛП3 и ЛП4. Пусть произвольно выбрали первой задачу ЛП4. Эта задача не имеет решения, и следовательно, является прозондированной. Следующей задачей выберем подзадачу ЛП3. Ее оптимальным решением является

$x_1 = 4{,}5, x_2 = 0, z = 22{,}5$. Нецелочисленное значение переменной $x_1 = 4{,}5$ порождает две ветви решения при $x_2 \le 4$ и $x_2 \ge 5$ и соответствующие им подзадачи ЛП5 и ЛП6. При этом

1) пространство решений ЛП5 = пространство решений ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$) +($x_1 \ge 4$).

2) пространство решений ЛП6 = пространство решений ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$) ($x_2 \ge 5$).

Теперь не рассмотрены лишь подзадачи ЛП1, ЛП5 и ЛП6. Подзадача ЛП6 прозондирована, так как не имеет допустимых решений. Подзадача ЛП5 имеет целочисленное решение $x_1 = 4, x_2 = 0, z = 20$ и следовательно, порождает нижнюю границу $z = 20$ оптимального значения целевой функции задачи ЦЛП. Решение подзадачи ЛП1 целочисленное $x_1 = 3, x_2 = 2, z = 23$. Следовательно нижнюю границу значений целевой функции полагаем равной 23. Так как все подзадачи прозондированы, оптимальным решением задачи ЦЛП является решение, соответствующее последней нижней границе, т.е. $x_1 = 3, x_2 = 2, z = 23$.

Последовательность решения подзадачи, показанная на рис. 5 (ЛП0, ЛП2, ЛП4, ЛП3, ЛП6, ЛП5, ЛП1) является наихудшей, тем не менее, она встречается на практике. Данный пример указывает на основную слабость метода ветвей и границ: как выбрать следующую подзадачу для исследования и как выбрать для нее переменную ветвления? Существуют эвристические соображения, позволяющие «угадать», какая из ветвей может привести к ухудшению решения задачи ЦЛП, при этом не существует строгой теории, которая всегда обеспечивала бы надежные результаты.

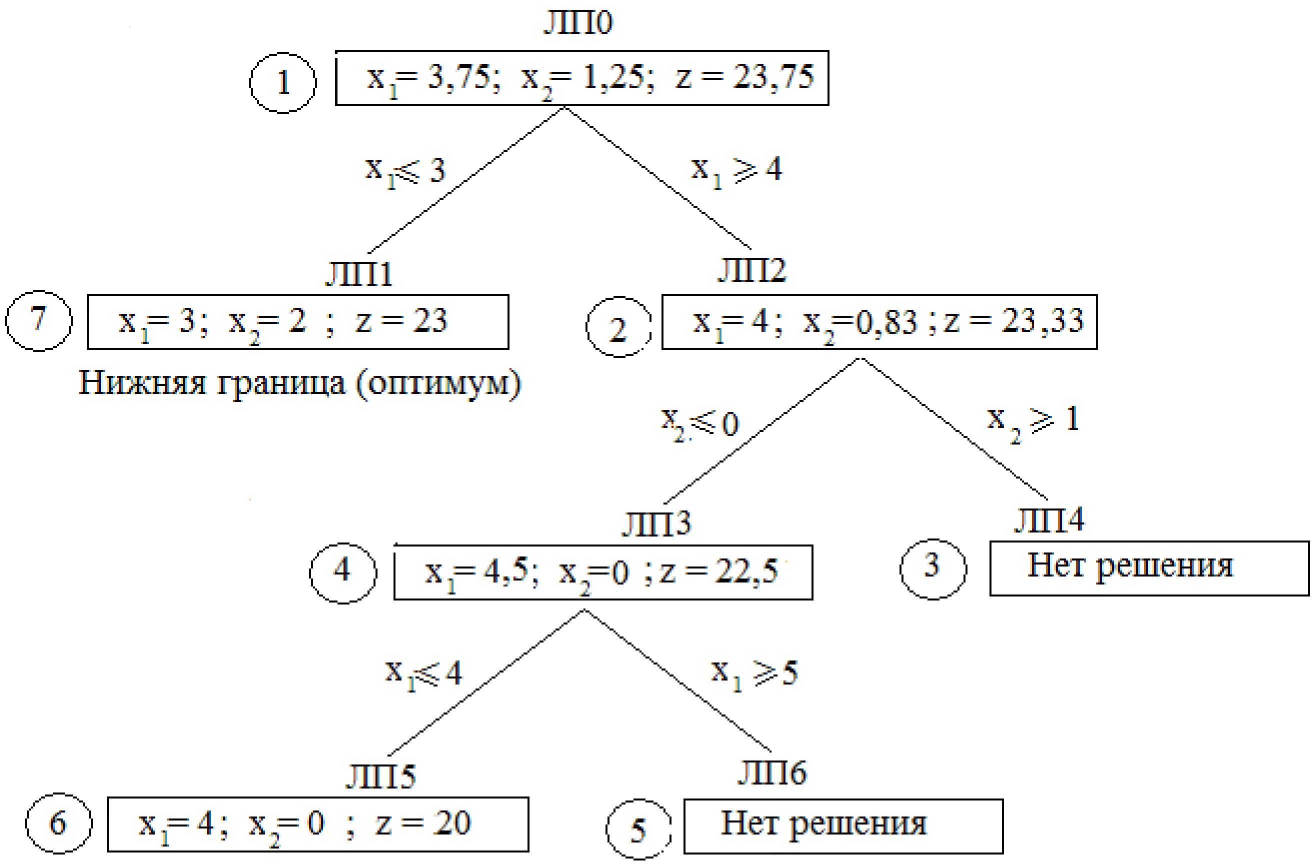


Рис. 5

Рассмотренная процедура применима для решения задач максимизации. Для решения задач минимизации в алгоритме необходимо заменить нижнюю границу верхней (начальное значение которой равно $z = +\infty$). Алгоритм метода ветвей и границ непосредственно распространяется на задачи частично-целочисленного ЛП, в которых лишь некоторые из переменных должны принимать целочисленное решение.

Вопросы для самопроверки:

1) Какие задачи в экономике требуют целочисленного решения?

2) Записать в общем виде математическую модель целочисленного программирования.

3) Как решить задачу целочисленного программирования графическим методом?

4) Какая геометрическая интерпретация решений целочисленной задачи на плоскости?

5) Привести алгоритм метода Гомори.

6) В чем суть процедуры формирования правильного отсечения?

6) Общая идея реализации метода ветвей и границ.

Тема 12. Методы нелинейного программирования

12.1. Общая постановка задачи выпуклого программирования

12.2. Графический метод

12.3. Метод множителей Лагранжа

12.4. Теорема Куна-Таккера

12.1. Общая постановка задачи выпуклого программирования

Нелинейное программирование применяется при оптимизации промышленного производства, управления товарными ресурсами, планировании обслуживания и ремонта оборудования и т.д.

Для задачи нелинейного программирования в отличие от линейных задач нет единого метода решения. В зависимости от вида целевой функции и системы ограничений разработаны специальные методы решения, к которым относятся методы множителей Лагранжа, квадратическое и выпуклое программирование, градиентные методы, приближенные методы решения, графический метод.

Задача математического программирования

$$
\begin{aligned}
&\max(\min) Z = f(x);\\
&\varphi_i(x)(\le,=,\ge)b_i \quad (i=\overline{1,m});\\
&\boldsymbol{x} \ge 0, \quad \boldsymbol{x} = (x_1,\dots,x_n),
\end{aligned}
$$

в которой либо целевая функция, либо ограничения, либо и то и другое нелинейные, называется нелинейной.

Нелинейные задачи составляют широкий класс настолько сложных задач, что до сих пор не разработаны совершенные общие методы, такие как симплекс-метод в линейном программировании, которые позволяли бы решать любые нелинейные задачи. Но, несмотря на отсутствие универсальных методов, разработаны способы решения отдельных специальных классов задач, и прежде всего задач с выпуклыми (вогнутыми) функциями $f(x)$ и $\varphi_i(x)$.

Определение. Функция $f(x)$, определенная на выпуклом множестве $\boldsymbol{X}$, называется выпуклой, если для любых точек x' и x'' из этого множества и любого $0 \le \lambda \le 1$ справедливо неравенство

$$f(\lambda x' + (1-\lambda)x'') \le \lambda f(x') + (1-\lambda)f(x'').$$

Если в данном соотношении при $0 < \lambda < 1$ и любых x', $x'' \in \boldsymbol{X}$ $(x' \neq x'')$ имеет место строгое неравенство, то $f(x)$ называется строго выпуклой.

Определение. Функция $f(x)$, определенная на выпуклом множестве $\boldsymbol{X}$, называется вогнутой, если для любых точек x' и x'' из этого множества и любого $0 \le \lambda \le 1$ справедливо неравенство

$$f(\lambda x' + (1-\lambda)x'') \ge \lambda f(x') + (1-\lambda) f(x'').$$

Если в данном соотношении при $0 < \lambda < 1$ и любых x', $x'' \in \boldsymbol{X}$ $(x' \neq x'')$ имеет место строгое неравенство, то $f(x)$ называется строго вогнутой.

Для решения задач нелинейного программирования важное значение имеют такие теоремы.

Теорема. Если $\varphi(x)$ - выпуклая функция при всех $\boldsymbol{x} \ge 0$, то будет выпуклым и множество решений системы $\varphi(x) \le b,\ \boldsymbol{x} \ge 0$.

Аналогичная теорема доказывается и для вогнутых функций. Доказывается, что выпуклая функция $f(x)$, определенная на выпуклом множестве $\boldsymbol{X}$, непрерывна в любой внутренней точке этого множества.

Теорема. Выпуклая функция $f(x)$, определенная на выпуклом множестве $\boldsymbol{X}$, достигает своего глобального минимума в каждой точке $\boldsymbol{x}$, в которой градиент функции обращается в нуль.

Теорема. Локальный минимум выпуклой функции $f(x)$, определенной на выпуклом множестве $\boldsymbol{X}$, совпадает с ее глобальным минимумом на этом множестве.

В теории выпуклого программирования в качестве основной обычно рассматривается задача минимизации выпуклой функции n переменных $Z = f(x)$ при ограничениях $\varphi_i(x) \le 0$ $(i = \overline{1,m})$, $\boldsymbol{x} \ge 0$, где функции $\varphi_i(x)$ предполагаются выпуклыми. Если $f(x)$ и $\varphi_i(x)$ являются вогнутыми функциями, то имеем задачу максимизации $f(x)$ при ограничениях $\varphi_i(x) \ge 0$ $(i = \overline{1,m})$, $\boldsymbol{x} \ge 0$.

12.2. Графический метод

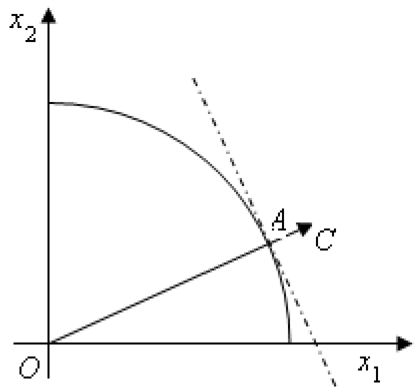
Задачи нелинейного программирования с двумя переменными, в которых их целевые функции и системы ограничений могут быть заданы в линейном и нелинейном виде, могут быть решены графически.

Задача с линейной целевой функцией и нелинейной системой ограничений

Пример. Найти глобальные экстремумы функции $F = 2x_1 + x_2$ при ограничениях:

$$\begin{cases} x_1^2 + x_2^2 \le 16 \\ x_{1,2} \ge 0 \end{cases}.$$

Решение. Область допустимых решений – часть окружности с радиусом 4, которая расположена в первой четверти. Линиями уровня целевой функции являются параллельные прямые с угловым коэффициентом , равным -2.



Решаем систему

$$\begin{cases} x_1^2 + x_2^2 = 16 \\ x_2 = \frac{1}{2} x_1 \end{cases} \Rightarrow x_1 = 8\sqrt{5}/5,\ x_2 = 4\sqrt{5}/5,\ F = 16\sqrt{5}/5 + 4\sqrt{5}/5 = 4\sqrt{5}.$$

Глобальный минимум, равный нулю, достигается в точке $O(0, 0)$, глобальный максимум, равный $4\sqrt{5}$ - в точке $A(8\sqrt{5}/5, 4\sqrt{5}/5)$.

12.3. Метод множителей Лагранжа

Постановка задачи.

Дана задача нелинейного программирования $F = f(x_1, x_2, ..., x_n) \to \max(\min)$ при ограничениях: $\varphi_i(x_1, x_2, ..., x_n) = 0, i = \overline{1,m}$.

Предположим, что функции $f(x_1, x_2, ..., x_n)$ и $\varphi_i(x_1, x_2, ..., x_n)$ непрерывны вместе со своими первыми частными производными.

Ограничения заданы в виде уравнений, поэтому для решения задачи используется метод отыскания условного экстремума функции нескольких переменных.

Для решения задачи составляют функцию Лагранжа

$L(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = f(x_1, x_2, ..., x_n) + \sum_{i=1}^{m} \lambda_i \varphi_i(x_1, x_2, ..., x_n)$, где λ_i - множители Лагранжа.

Затем определяются частные производные

$$\frac{\partial L}{\partial x_j}, \frac{\partial L}{\partial \lambda_i}, i = \overline{1, m},\ j = \overline{1, n}.$$

Приравнивая к нулю частные производные, получим систему

$$\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial \varphi_i}{\partial x_j} = 0,$$
$$\frac{\partial L}{\partial \lambda_i} = g_i(x_1, x_2, ..., x_n).$$

Решая систему, получим множество точек, в которых целевая функция L может иметь экстремальные значения. Следует отметить, что условия рассмотренной системы являются необходимыми, но недостаточными. Поэтому не всякое полученное решение определяет точку экстремума целевой функции. Применение метода бывает оправданным, когда заранее предполагается существование глобального экстремума, совпадающего с единственным локальным максимумом или минимумом целевой функции.

Пример. Мукомольный комбинат реализует муку двумя способами: в розницу через магазин и оптом через торговых агентов. При продаже x_1 кг муки через магазин расходы на реализацию составляют x_1^2 ден. ед., а при продаже x_2 кг муки посредством торговых агентов - x_2^2 ден. ед.

Определить, сколько килограммов муки следует продавать каждым способом, чтобы затраты на реализацию были минимальными, если в сутки выделяется для продажи 5 000 кг муки.

Решение. Составим математическую модель задачи.

Найдем минимум суммарных расходов $L = x_1^2 + x_2^2$ при ограничениях $x_1 + x_2 = 5000,\quad x_{1,2} \geq 0$. Для расчета модели используем метод множителей Лагранжа. Составим функцию Лагранжа $F(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 5000)$.

Найдем частные производные функции F по x_1, x_2, λ, приравняем их к нулю, получим систему уравнений

$$\begin{cases} 2x_1 + \lambda = 0, \\ 2x_2 + \lambda = 0, \\ x_1 + x_2 - 5000 = 0, \end{cases}$$

Откуда $\lambda = -5000,\quad x_1 = 2500, x_2 = 2500, L = 12500000$ ден.ед. Придавая x_1 значения больше и меньше 2500, находим L и из определения экстремума функции получаем, что L при $x_1 = x_2 = 2500$ достигает минимума.

Таким образом, для получения минимальных расходов необходимо расходовать в сутки через магазин и торговых агентов по 2500 кг муки, при этом расходы на реализацию составят 12500000 ден. ед.

12.4. Теорема Куна-Таккера

В теории нелинейного программирования центральное место занимает теорема Куна-Таккера, обобщающая классический метод множителей Лагранжа на случай, когда в нелинейной задаче, помимо ограничений-равенств, содержатся также и неравенства. В частности, для задачи выпуклого программирования: минимизировать $Z = f(x)$ при ограничениях $\varphi_i(x) \le 0 \quad (i = \overline{1,m}),\ \boldsymbol{x} \ge 0$, где все функции $f(x)$ и $\varphi_i(x)$ выпуклые, теорема Куна-Таккера устанавливает связь между оптимальным решением задачи и Седловой точкой функции Лагранжа для этой задачи:

$$L(x, \overline{\lambda}) = f(x) + \sum_{i=1}^{m} \lambda_i \varphi_i(x).$$

Точка $(x^*, \overline{\lambda}^*)$ называется седловой точкой функции $L(x, \overline{\lambda}) = f(x) + \sum_{i=1}^{m} \lambda_i \varphi_i(x)$, если n-мерная точка x^* является точкой минимума функции $L(x, \overline{\lambda}^*)$, а m-мерная точка $\overline{\lambda}^*$ - точкой максимума функции $L(x^*, \overline{\lambda})$, так что для всех x и $\overline{\lambda}$ выполняется неравенство $L(x^*, \overline{\lambda}) \le L\left(x^*, \overline{\lambda}^*\right) \le L\left(x, \overline{\lambda}^*\right)$.

Теорема Куна-Таккера. Предположим, что существует вектор $\boldsymbol{x} \ge 0$, такой что $\varphi_i(x) < 0 \quad (i = \overline{1,m})$. Тогда необходимым и достаточным условием оптимальности вектора $\boldsymbol{x}^*$, принадлежащего допустимой области, является существование такого вектора $\overline{\lambda}^*$, что для всех $\boldsymbol{x} \ge 0$ и $\overline{\lambda}$ имеет место неравенство

$$L(x^*, \overline{\lambda}) \le L\left(x^*, \overline{\lambda}^*\right) \le L\left(x, \overline{\lambda}^*\right).$$

Если функции $f(x)$ и $\varphi_i(x)$ являются дифференцируемыми, то данное неравенство, где $x^* \ge 0$, $\bar{\lambda}^* \ge 0$, эквивалентны следующим «локальным» условиям Куна-Таккера:

$$\frac{\partial L(x^*, \bar{\lambda}^*)}{\partial x} \ge 0,$$
$$\frac{x^* \partial L(x^*, \bar{\lambda}^*)}{\partial x} = 0,$$
$$x^* \ge 0,$$
$$\frac{\partial L(x^*, \bar{\lambda}^*)}{\partial \bar{\lambda}} \le 0,$$
$$\frac{\bar{\lambda}^* \partial L(x^*, \bar{\lambda}^*)}{\partial \bar{\lambda}} = 0$$
$$\bar{\lambda}^* \ge 0.$$

Вопросы для самопроверки:

1) В чем состоит отличие задач нелинейного программирования от линейного?
2) Привести общую постановку задачи выпуклого программирования.
3) Как задача нелинейного программирования решается графически?
4) В чем состоит идея метода множителей Лагранжа?
5) В чем суть теоремы Куна-Таккера?

Тема 13. Квадратичное программирование

13.1. Основные понятия квадратичного программирования

13.2. Алгоритм решения задачи квадратичного программирования

13.1. Основные понятия квадратичного программирования

В качестве основной задача в квадратичном программировании принимается задача:

$$f(x) = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} x_i x_j \to \max(\min) \qquad (13.1)$$

при ограничениях:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i \ , \\ (i = \overline{1, n}), \\ x_j \ge 0, \\ (j = \overline{1, n}). \end{cases} \qquad (13.2)$$

Для решения задачи необходимо привести определения.

Квадратичной формой относительно переменных $x_1, ..., x_n$ называется числовая функция от этих переменных, имеющая вид:

$$F = d_{11}x_1x_1 + d_{12}x_1x_2 + d_{13}x_1x_3 + ... + d_{21}x_2x_1 + d_{22}x_2x_2 + ... = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_ix_j \ .$$

Матрица $D = (d_{ij})$ будет отрицательно определенной в задаче максимизации и положительно определенной – в задаче минимизации. Это означает, что функция $f(x)$ является строго выпуклой по переменным x в случае задачи минимизации и строго вогнутой – в задаче максимизации. Ограничения в этой задаче предполагаются линейными, что гарантирует выпуклость области допустимых решений.

Квадратичная форма F называется положительно (отрицательно)-полуопределенной, если $F(X) \ge 0$ ($F(X) \le 0$) для любого набора значений переменных $X = (x_1, ..., x_n)$ и существует такой набор переменных $X' = (x'_1, ..., x'_n)$, где не все значения переменных одновременно равны нулю $F(X') = 0$.

Теорема. Квадратичная форма является выпуклой функцией, если она положительно-полуопределенная, и вогнутой функцией, если она отрицательно-полуопределенная.

В случае, если целевая функция не минимизируется, а максимизируется или, если в некоторых ограничениях имеется знак ≥, то такие задачи всегда можно привести к основной форме (13.1) − (13.2). Отличием задачи квадратичного программирования от задачи линейного программирования есть то, что в последней оптимальное решение находится в вершине многогранника, а в задачи (13.1) − (13.2) оптимальное решение может быть как на границе, так и внутри многогранника решений.

Для решения задачи (13.1) − (13.2) составим локальные условия Куна – Таккера, являющиеся необходимыми и достаточными условиями оптимальности решения:

$$\frac{\partial L(x^*,\overline{\lambda}^*)}{\partial x} \geq 0,$$
$$\frac{x^* \partial L(x^*,\overline{\lambda}^*)}{\partial x} = 0,$$
$$x^* \geq 0,$$
$$\frac{\partial L(x^*,\overline{\lambda}^*)}{\partial \overline{\lambda}} \leq 0,$$
$$\frac{\overline{\lambda}^* \partial L(x^*,\overline{\lambda}^*)}{\partial \overline{\lambda}} = 0$$
$$\overline{\lambda}^* \geq 0.$$

Функция Лагранжа в данном случае имеет вид:

$$L = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} x_i x_j + \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right).$$

Найдем частные производные данной функции:

$$\frac{\partial L}{\partial x_j} = c_j + 2\sum_{i=1}^{n} d_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} \quad (j = \overline{1, n}),$$

$$\frac{\partial L}{\partial \lambda_i} = \sum_{j=1}^{n} a_{ij} x_j - b_i \quad (i = \overline{1, m}).$$

Обозначим

$$c_j + 2\sum_{i=1}^{n} d_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} = v_j,$$

$$\sum_{j=1}^{n} a_{ij} x_j - b_i = -y_i \quad .$$

С учетом этих обозначений «локальные» условия Куна - Таккера примут вид:

$$c_j + 2\sum_{i=1}^{n} d_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} = v_j \geq 0,$$
$$x_j v_j = 0, \quad x_j \geq 0,$$
$$\sum_{j=1}^{n} a_{ij} x_j - b_i = -y_i \leq 0,$$
$$\lambda_i (-y_i) = 0, \quad \lambda_i \geq 0.$$

Объединяя эти соотношения, получаем:

$$\sum_{j=1}^{n} a_{ij}x_j + y_i = b_i \ , \qquad (13.3)$$

$$2\sum_{i=1}^{n} d_{ij}x_i - v_j + \sum_{i=1}^{m} \lambda_i a_{ij} = -c_j , \qquad (13.4)$$

$$x_j \geq 0,\ v_j \geq 0,\ y_i \geq 0,\ \lambda_i \geq 0,$$

$$\sum_{j=1}^{n} x_j v_j + \sum_{i=1}^{m} y_i \lambda_i = 0 . \qquad (13.5)$$

Получили систему $n+m$ линейных уравнений с $2(n+m)$ неизвестными $x_1, ..., x_n, v_1, ..., v_n, y_1, ..., y_m, \lambda_1, ..., \lambda_m$.

Поскольку функция $f(x)$ строго вогнута, а область допустимых решений представляет собой выпуклое множество, допустимое решение, удовлетворяющее всем этим условиям, должно быть единственным и оптимальным.

Следовательно, в квадратичном программировании с помощью теоремы Куна – Таккера нелинейную задачу можно сводить к решению системы линейных алгебраических уравнений при соблюдении дополнительного условия $\sum_{j=1}^{n} x_j v_j + \sum_{i=1}^{m} y_i \lambda_i = 0$, согласного которого: из каждых двух ограниченных по знаку переменных x_j и v_j (соответственно y_i и λ_i) хотя бы одна равнялась нулю. Другими словами, не все решения $X = (x_1, ..., x_n, v_1, ..., v_n, y_1, ..., y_m, \lambda_1, ..., \lambda_m)$ системы (13.3), (13.4) удовлетворяют условию (13.5), а только те, в которых по крайней мере $n+m$ компонент равны нулю, т.е. столько, сколько уравнений в системе и искать решение следует базисных решений. Решение завершается обращением в нуль всех искусственных переменных только в том случае, когда исходная задача имеет непустое множество допустимых решений.

13.2. Алгоритм решения задачи квадратичного программирования

Решение задачи квадратичного программирования можно найти с помощью метода искусственного базиса, примененного для нахождения максимального значения функции $F = -\sum_{i=1} M y_i$ при условиях (13.3) − (13.5), где y_i − искусственные переменные, введенные в уравнения (13.3), (13.4).

Алгоритм решения задачи квадратичного программирования следующий:

1) составляют функцию Лагранжа;
2) записывают в виде выражений (13.3) − (13.5) необходимые и достаточные условия существования седловой точки для функции Лагранжа;
3) используя метод искусственного базиса, либо устанавливают отсутствие седловой точки для функции Лагранжа, либо находят ее координаты;
4) записывают оптимальное решение исходной задачи и находят значение целевой функции.

Помимо общих методов выпуклого программирования, специально для задачи выпуклого квадратичного программирования разработано большое количество различных численных методов, в том числе и конечных. Наиболее употребительные конечные методы: симплексный метод Била, метод сопряженных градиентов, симплексный метод Вульфа. Однако эффективность того или иного метода весьма существенно зависит от специфики решаемой задачи и имеющихся в распоряжении вычислительных средств.

Пример. Найти максимальное значение функции $f = 2x_1 + 4x_2 - x_1^2 - 2x_2^2$

При условиях

$$\begin{cases} x_1 + 2x_2 \le 8, \\ 2x_1 - x_2 \le 12, \\ x_1, x_2 \ge 0. \end{cases}$$

Решение. Функция f является вогнутой, поскольку представляет собой сумму линейной функции $f_1 = 2x_1 + 4x_2$, которую можно рассматривать как вогнутую и квадратичной формы $f_2 = -x_1^2 - 2x_2^2$, которая является отрицательно-определенной и, следовательно, также вогнутой. Система ограничений задачи включает только линейные неравенства. Следовательно, можно воспользоваться теоремой Куна – Таккера.

Составим функцию Лагранжа:

$$L = 2x_1 + 4x_2 - x_1^2 - 2x_2^2 + \lambda_1 (8 - x_1 - 2x_2) + \lambda_2 (12 - 2x_1 + x_2).$$

Запишем необходимые и достаточные условия существования седловой точки функции:

$$\frac{\partial L}{\partial x_1} = 2 - 2x_1 - \lambda_1 - 2\lambda_2 \le 0,$$

$$\frac{\partial L}{\partial x_2} = 4 - 4x_2 - 2\lambda_1 + \lambda_2 \le 0,$$

$$\frac{\partial L}{\partial \lambda_1} = 8 - x_1 - 2x_2 \ge 0,$$

$$\frac{\partial L}{\partial \lambda_2} = 12 - 2x_1 + x_2 \ge 0,$$

$$x_1 \frac{\partial L}{\partial x_1} = x_1 (2 - 2x_1 - \lambda_1 - 2\lambda_2) = 0,$$

$$x_2 \frac{\partial L}{\partial x_2} = x_2 (4 - 4x_2 - 2\lambda_1 + \lambda_2) = 0,$$

$$\lambda_1 \frac{\partial L}{\partial \lambda_1} = \lambda_1 (8 - x_1 - 2x_2) = 0,$$

$$\lambda_2 \frac{\partial L}{\partial \lambda_2} = \lambda_2 (12 - 2x_1 + x_2) = 0,$$

$$x_1, x_2, \lambda_1, \lambda_2 \ge 0.$$

Таким образом, имеем систему неравенств:

$$2x_1 + \lambda_1 + 2\lambda_2 \ge 2,$$

$$4x_2 + 2\lambda_1 - \lambda_2 \ge 4,$$

$$x_1 + 2x_2 \le 8,$$

$$2x_1 - x_2 \le 12.$$

Преобразуем систему неравенств в равенства, введя балансовые переменные:

$$2x_1 + \lambda_1 + 2\lambda_2 - v_1 = 2,$$

$$4x_2 + 2\lambda_1 - \lambda_2 - v_2 = 4, \qquad (13.6)$$

$$x_1 + 2x_2 + w_1 = 8,$$

$$2x_1 - x_2 + w_2 = 12.$$

Учитывая равенства (13.5), запишем:

$$v_1 x_1 = 0, v_2 x_2 = 0, w_1 \lambda_1 = 0, w_2 \lambda_2 = 0.$$

Для нахождения базисного решения системы линейных уравнений (13.6) следует применить метод искусственного базиса. В первое и во второе уравнения системы (13.6) добавим искусственные неотрицательные переменные z_1, z_2. Получим расширенную задачу:

$$F = -M z_1 - M z_2$$

при условиях:

$$\begin{cases} 2x_1 + \lambda_1 + 2\lambda_2 - v_1 + z_1 = 2, \\ 4x_2 + 2\lambda_1 - \lambda_2 - v_2 + z_2 = 4, \\ x_1 + 2x_2 + w_1 = 8, \\ 2x_1 - x_2 + w_2 = 12, \\ x_1, x_2, \lambda_1, \lambda_2, v_1, v_2, w_1, w_2, z_1, z_2 \ge 0. \end{cases}$$

Решим задачу симплекс-методом.

				0	0	0	0	0	0	0	0	$-M$	$-M$
N	Базис	$C_б$	A_0	A_{x_1}	A_{x_2}	A_{λ_1}	A_{λ_2}	A_{v_1}	A_{v_2}	A_{w_1}	A_{w_2}	A_{z_1}	A_{z_2}
1	A_{z_1}	$-M$	2	2	0	1	2	-1	0	0	0	1	0
2	A_{z_2}	$-M$	4	0	**4**	2	-1	0	-1	0	0	0	1
3	A_{w_1}	0	8	1	2	0	0	0	0	1	0	0	0
4	A_{w_2}	0	12	2	-1	0	0	0	0	0	1	0	0
		Δ_1	$-6M$	$-2M$	$-4M$	$-3M$	$-M$	M	M	0	0	0	0
5	A_{z_1}	$-M$	2	**2**	0	1	2	-1	0	0	0	1	0
6	A_{x_2}	0	1	0	**1**	$\frac{1}{2}$	$-\frac{1}{4}$	0	$-\frac{1}{4}$	0	0	0	$\frac{1}{4}$
7	A_{w_1}	0	6	1	0	-1	$\frac{1}{2}$	0	$\frac{1}{2}$	1	0	0	$-\frac{1}{2}$
8	A_{w_2}	0	13	2	0	$\frac{1}{2}$	$-\frac{1}{4}$	0	$-\frac{1}{4}$	0	1	0	$\frac{1}{4}$
		Δ_2	$-2M$	$-2M$	0	$-M$	$-2M$	M	0	0	0	0	M
9	A_{x_1}	0	1	**1**	0	$\frac{1}{2}$	1	$-\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0
10	A_{x_2}	0	1	0	**1**	$\frac{1}{2}$	$-\frac{1}{4}$	0	$-\frac{1}{4}$	0	0	0	$\frac{1}{4}$
11	A_{w_1}	0	5	0	0	$-\frac{3}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0	$-\frac{1}{2}$	$-\frac{1}{2}$
12	A_{w_2}	0	11	0	0	$-\frac{1}{2}$	$-\frac{9}{4}$	1	$-\frac{1}{4}$	0	1	-1	$\frac{1}{4}$
		Δ_3	0	0	0	0	0	0	0	0	0	M	M

Допустимое базисное решение:
$x_1 = 1, x_2 = 1, \lambda_1 = 0, \lambda_2 = 0, v_1 = 0, v_2 = 0, w_1 = 5, w_2 = 11, z_1 = 0, z_2 = 0$.
Поскольку выполняются условия $v_1 x_1 = 0, v_2 x_2 = 0, w_1 \lambda_1 = 0, w_2 \lambda_2 = 0$, то $X^* = (1,1)$ является седловой точкой функции Лагранжа для исходной задачи. Следовательно, $X^* = (1,1)$ - оптимальный план исходной задачи , $f(X^*) = 3$.

Вопросы для самопроверки:

1) Привести математическую модель основной задача в квадратичном программировании.
2) В чем отличие задачи квадратичного программирования от задачи линейного программирования?
3) В каких целях в квадратичном программировании используется теорема Куна – Таккера?
4) Назвать этапы алгоритма решения задачи квадратичного программирования.

Тема 14. Теория игр. Основные методы их решения и анализа

14.1. Основные понятия теории игр

14.2. Решения матричной игры в чистых стратегиях

14.3. Решение матричной игры в смешанных стратегиях

14.4. Решение матричной игры графическим методом

14.1. Основные понятия теории игр

В практической деятельности часто необходимо согласовать действия фирм, предприятий, объединений, министерств и других участников проектов в случаях, когда их интересы не совпадают. В этих ситуациях теория игр позволяет найти лучшее решение в поведении всех участников, обязанных согласовывать действия при столкновении интересов. Теория игр широко проникла в практику экономических решений и исследований. Так, можно определить научно обоснованные уровни снижения розничных цен и оптимальный уровень товарных запасов, решить задачу планирования порядка организации источников полезных ресурсов в стране и т.д.

Теорию игр определяют как раздел математики, которая изучает конфликтные ситуации. Это значит, что можно сформулировать оптимальные правила поведения каждой стороны, которая принимает участие в решении конфликтной ситуации. В экономике оказалось недостаточно аппарата математического анализа, связанного с определением экстремумов функции. Теорию игр следует рассматривать как отдельный раздел оптимизационного подхода, позволяющий решать новые задачи в принятии решения.

Игра – это упрощенная формализованная модель реальной конфликтной ситуации, отличная от реального конфликта тем, что ведется по определенным правилам. Другими словами, игра – это совокупность правил, определяющих возможные действия (чистые стратегии) участников игры.

Суть игры состоит в том, что каждый из участников принимает решения в развивающейся конфликтной ситуации, которые как он понимает, могут обеспечить ему самый лучший исход. Исход игры – это значение некоторой функции, называемой функцией выигрыша (платежной функцией). Эта функция задается либо таблицей, либо аналитическим выражением. Если сумма выигрышей игроков равна нулю, то игру называют игрой с нулевой суммой. Если в игре участвуют два игрока, то ее называют парной.

Величина выигрыша зависит от стратегии, которую использует игрок. Стратегия – это совокупность правил, однозначно определяющая последовательность действий игрока в каждой конкретной ситуации, реализующейся в процессе игры. Оптимальной называется стратегия, которая при многократном повторении игры обеспечивает данному игроку максимально возможный средний выигрыш. Всякая игра состоит из партий. Партией называют каждый вариант игры определенным образом. В свою очередь, в партии игроки осуществляют конкретные ходы. Ход – это выбор и реализация игроком одного из допустимых вариантов поведения.

В зависимости от количества стратегий игры делятся на конечные и бесконечные. По характеру выигрышей игры делятся на игры с нулевой суммой и игры с ненулевой суммой. В первых – общий капитал игроков не меняется, а лишь перераспределяется в ходе игры, в связи с чем сумма выигрышей равна нолю (проигрыш принимается как отрицательный выигрыш). В последующем мы будем рассматривать только парную матричную игру с нулевой суммой.

Игра, в которой участники стремятся достичь для себя самого лучшего результата, осознанно выбирая допустимыми правилами игры способы действий, называют стратегической. Однако в экономике часто приходится моделировать (формализовать) ситуации, в которой один из участников безразличен к результату игры. Такая игра называется игрой с природой, понимая под природой всю совокупность внешних обстоятельств, в которых осознанному игроку приходится принимать решение. Например, определение объема производства выпуска сезонной продукции в надежде наиболее выгодного дня ее реализации для уровня спроса; формирование пакета ценных бумаг в расчете на высокие дивиденды и т.д. Здесь в роли игрока выступает в первом случае уровень спроса, во втором – размер ожидаемой прибыли. В игре с природой мера неопределенности осознанного игрока возрастает.

Рассмотрим игру, в которой у каждого из двух игроков A и B имеется конечное число возможных действий, т.е. чистых стратегий. Допустим, что игрок A имеет в своем распоряжении m чистых стратегий, которые обозначим $A_1, ..., A_m$, а игрок B имеет n чистых стратегий $B_1, ..., B_n$. Для того, чтобы игра была полностью определена, необходимо указать правило, по которому каждой паре чистых стратегий $(A_i; B_j)$ ставится в соответствие число a_{ij} – выигрыш игрока A за счет игрока B или проигрыш игрока B. Мы рассматриваем игру с нулевой суммой. При $a_{ij} < 0$ игрок A выплачивает игроку B сумму $|a_{ij}|$. Если игра состоит только с точных ходов, то выбор пары чистых стратегий $(A_i; B_j)$ единственным способом определяет исход игры. Если же в течении игры используются случайные ходы, то исход игры определяется средним значением выигрыша (его математическим ожиданием).

Пусть известны значения a_{ij} для каждой пары $(A_i; B_j)$ чистых стратегий, следовательно, можно составить матрицу игры, т.е. платежную матрицу, которая есть табличной записью функции выигрыша.

A_i	B_j				α_i
	B_1	B_2	$\dots$	B_n	
A_1 $\dots$ A_m	a_{11} $\dots$ a_{m1}	a_{12} $\dots$ a_{m2}	$\dots$ $\dots$ $\dots$	a_{1n} $\dots$ a_{mn}	α_1 $\dots$ α_m
β_j	β_1	β_2	$\dots$	β_n	

Отдельная партия в такой игре реализуется так: игрок A выбирает одну из строк платежной матрицы (одну из своих чистых стратегий), не зная результата его выбора, игрок B выбирает один из столбцов (свою чистую стратегию). Элемент матрицы, стоящий на пересечении выбранных строки и столбца определяет выигрыш игрока A (проигрыш игрока B).

14.2. Решения матричной игры в чистых стратегиях

Целью участников всякой матричной игры является выбор наиболее выгодных стратегий, обеспечивающих игроку A максимальный выигрыш, а игроку B − минимальный проигрыш.

При поиске оптимальных стратегий игроки опираются на основополагающий принцип теории игр – принцип осторожности, согласно которому каждый игрок, считая противника по игре очень разумным, выбирает свои действия в предположении, что противник ни в коем случае не упустит ни единой возможности использовать любую его ошибку, любой просчет в своих интересах. Поэтому игроки должны быть очень внимательны при выборе каждой своей чистой стратегии.

Пусть игроку A следует сделать свой выбор. Анализируя платежную матрицу (табл.1), он для каждой чистой стратегии A_i $(i = \overline{1, m})$ найдет минимальное значение α_i возможного выигрышу: $\alpha_i = \min\limits_j \{a_{ij}\}$ $(i = \overline{1, m})$, а потом со всех α_i выберет наибольшее $\alpha = \max\limits_i \{a_i\}$, и примет соответствующую ему чистую стратегию A_i^0. Это и будет преобладающая (гарантирующая) в данных условиях стратегия игрока A. Ее называют максиминной, поскольку она соответствует величине

$$\alpha = \max_i \min_j \{a_{ij}\}$$

Число α называется нижней чистой ценой игры (максиминной). Оно показывает, какой минимальный выигрыш может получить игрок A при любых действиях игрока B, если правильно использует свои чистые стратегии.

В свою очередь, игрок B, стремясь минимизировать проигрыш, при выборе преобладающей стратегии использует принцип осторожности таким образом: сначала для каждой чистой стратеги B_j $(j = \overline{1, n})$ он определяет максимально возможный проигрыш $\beta_j = \max\limits_i \{a_{ij}\}$, де $j = \overline{1, n}$ (табл.1), а затем среди β_j выберет минимальное значение $\beta = \min\limits_j \{\beta_j\}$, которому и будет отвечать искомая чистая стратегия B_j^0. Ее называют минимаксной, поскольку она отвечает величине

$$\beta = \min_j \max_i \{a_{ij}\}.$$

Число β, называется верхней чистой ценой игры (минимаксной). Оно показывает, какой максимальный проигрыш может бать в игрока B при правильном выборе им своих чистых стратегий не зависимо от действий игрока A.

Следовательно, правильно используя чистые стратегии, игрок A обеспечивает себе выигрыш не меньше α, а игрок B, правильно используя свои чистые стратеги, не позволит игроку A выиграть больше, чем β.

Теорема: В матричной игре нижняя чистая цена игры не превышает верхней чистой цены игры, т.е. $\alpha \le \beta$ (без доказательства).

Если в матричной игре нижняя и верхняя чистые цены совпадают, т.е. $\alpha = \beta$, то игра имеет седловую точку в чистых стратегиях и чистую цену игры $v = \alpha = \beta$.

Обозначим через i_* и j_* номера чистых стратегий, при которых имеет место равенство $\alpha = \beta$. Пара чистых стратегий $(A_{i*}; B_{j*})$ игроков A и B, при которых достигается это равенство, называется седловой точкой матричной игры, а элемент $a_{i_* j_*}$ матрицы, который стоит на пересечении i_*-ой строки и j_*-ого столбца, является седловым элементом платежной матрицы:

$$\max_i \min_j \{a_{ij}\} = \min_j \max_i \{a_{ij}\} = a_{i_* j_*} .$$

Пример. Имеются три возможных стратегии производства продукции:

$S_1 = 986\,200{,}50$ тыс. грн., $S_2 = 1488\,973$ тыс. грн., $S_3 = 1967\,962{,}1$ тыс. грн.

В зависимости от изменений рыночной конъюнктуры в связи с имеющимися возможностями реализации рассчитаны варианты среднегодовой прибыли (таблица).

Таблица 14.1

Матрица платежеспособного спроса

Объем производства (оптовые закупки)	**Размер прибыли (a_{ij}) в зависимости от возможных колебаний спроса**				$\alpha_i = \min\{a_{ij}\}$
	489 876,17	986 200,50	1488 973	1967 962,1	
S_1=986 200,50	49 310,03	197 240,1	197 240,1	197 240,1	49 310,03
S_2=1488 973	-60	148 897,3	297 794,6	297 794,6	-60
S_3=1967 962,1	-1140	98 398,11	196 796,24	393 592,42	-1140

$\beta_j = \max \{a_{ij}\}$	49 310,03	197 240,1	297 794,6	393 592,42	

Решение. Анализ игры начнем с позиции максимина:

$$\alpha_1 = \min_j \{a_{ij}\} = \min \{49310{,}01;\ 197240{,}10;\ 197240{,}10;\ 197240{,}10\} =$$
$$= 49310{,}33 \text{ тис. грн.}$$

$$\max_i \{\alpha_i\} = \max_i \min_j \{a_{ij}\} = \max\{49310{,}03;\ -60;\ -1140\} = 49310{,}03 \text{ тис. грн.}$$

Стратегия S_1 называется максиминной, и это есть нижняя цена игры.

Аналогично будем рассуждать со стороны «природы»:

$$\beta_1 = \max_i \{49310{,}03;\ -60;\ -1140\} = 49310{,}03 \text{ тис. грн}.$$

$$\beta = \min_j \{\beta_j\} = \min_j \{49310{,}03;\ 197240{,}1;\ 297794{,}6;\ 393592{,}42\} =$$
$$= 49310{,}03 \text{ тис. грн.}$$

Эта величина называется верхней ценой игры. Седловая точка равна $v = \alpha_1 = \beta_1 = 49310{,}03$ (табл.). Таким образом, необходимо выбрать первую стратегию с объемами производства 986200,50 тыс. грн.

14.3. Решение матричной игры в смешанных стратегиях

Если матричная игра не имеет седловой точки, то решение игры затрудняется. В этих играх $\alpha < \beta$. Решение находят, применяя смешанные стратегии.

Смешанной стратегией первого (второго) игрока называется вектор $p = (p_1; p_2; \ldots; p_m)$, где $p_i \ge 0$ $(i = \overline{1,m})$ и $\sum_{i=1}^{m} p_i = 1$ ($q = (q_1; q_2; \ldots; q_m)$, где $q_j \ge 0$ $(j = \overline{1,n})$ и $\sum_{j=1}^{n} q_j = 1$).

Вектор $\boldsymbol{p}(\boldsymbol{q})$ означает вероятность применения i-й чистой стратегии первым игроком (j-й чистой стратегии вторым игроком).

Поскольку игроки выбирают свои чистые стратегии случайно и независимо друг от друга, игра имеет случайный характер и случайной становится величина выигрыша (проигрыша). В таком случае средняя величина выигрыша (проигрыша) – математическое ожидание – является функцией смешанных стратегий $\boldsymbol{p}, \boldsymbol{q}$:

$$f(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} p_i q_j \ .$$

Функция $f(\boldsymbol{p},\boldsymbol{q})$ называется платежной функцией игры с матрицей $(a_{ij})_{m\times n}$.

Стратеги $p^*=(p_1^*;p_2^*;...;p_m^*)$, $q^*=(q_1^*;q_2^*;...;q_m^*)$ называется оптимальными, если для произвольных стратегий $p=(p_1;p_2;...;p_m)$, $q=(q_1;q_2;...;q_m)$ выполняется условие:

$$f(\boldsymbol{p},\boldsymbol{q}^*)\le f(\boldsymbol{p}^*,\boldsymbol{q}^*)\le f(\boldsymbol{p}^*,\boldsymbol{q}).$$

Использование в игре оптимальных смешанных стратегий обеспечивает первому игроку выигрыш, не меньший, чем при использовании им любой другой стратегии $\boldsymbol{p}$, второму игроку - проигрыш, не больший , чем при использовании им любой другой стратегии $\boldsymbol{q}$.

Совокупность оптимальных стратегий и цены игры составляет решение игры.

Значение платежной функции при оптимальных стратегиях определяет цену игры ν, т.е. $f(\boldsymbol{p}^*,\boldsymbol{q}^*)=\nu$.

Теорема. В смешанных стратегиях любая конечная матричная игра имеет седловую точку.

Теорема. Для того, чтобы смешанные стратегии $p^*=(p_1^*;p_2^*;...;p_m^*)$ и $q^*=(q_1^*;q_2^*;...;q_m^*)$ были оптимальными для игроков A и B в игре с матрицей $(a_{ij})_{m\times n}$ и выигрышем ν, необходимо и достаточно выполнение неравенств:

$$\sum_{i=1}^{m}a_{ij}p_i^*\ge\nu \qquad (j=\overline{1,n}),$$

$$\sum_{j=1}^{n}a_{ij}q_j^*\le\nu \qquad (i=\overline{1,m}).$$

Теорема. Если один из игроков придерживается своей оптимальной смешанной стратегии, то его выигрыш остается неизменным и равным цене игры независимо от того, какую стратегию применяет другой игрок, если только тот не выходит за пределы своих активных стратегий.

14.4. Решение матричной игры графическим методом

Наиболее простая матричная игра – это игра, в которой каждый из игроков имеет две стратегии. Матрица $\mathbf{A}$ имеет вид:

$$\mathbf{A}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}.$$

Если седловой точки нет, то решением игры находится в смешанных стратегиях, а именно $\mathbf{X} = (x_1, x_2)$, $\mathbf{Y} = (y_1, y_2)$. Это есть вероятности, с которыми применяют игроки свои стратегии.

Запишем систему уравнений

$$\begin{cases} a_{11}x_1 + a_{21}x_2 = v \\ a_{12}x_1 + a_{22}x_2 = v \end{cases}.$$

С учетом $x_1 + x_2 = 1$, находим

$$x_1 = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}, \quad x_2 = \frac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}. \qquad (1)$$

Цена игры:

$$v = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}. \qquad (2)$$

Составляя аналогичную систему уравнений, находим оптимальную стратегію для игрока B:

$$y_1 = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}, \quad y_2 = \frac{a_{11} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}. \qquad (3)$$

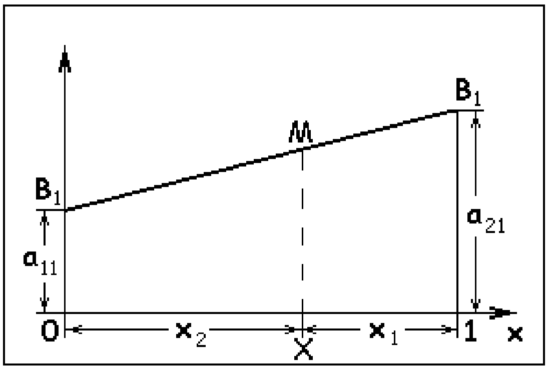
Пример. Решить игру, заданной платежной матрицей:

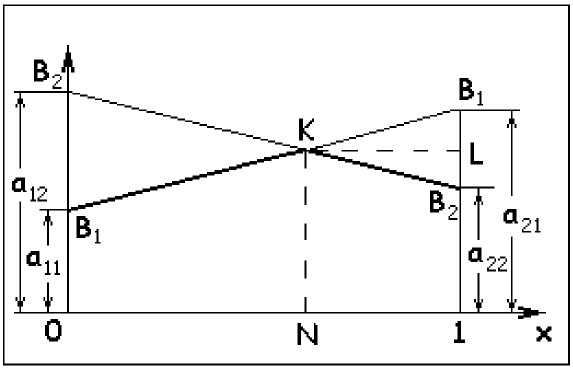
$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}.$$

Решение. Имеем, что $\alpha = 1$, $\beta = 2$, матрица не имеет седловой точки. По формулам находим оптимальные стратегии и цену игры:

$$\mathbf{X} = \left(\frac{1}{3}; \frac{2}{3}\right), \ \mathbf{Y} = \left(\frac{2}{3}; \frac{1}{3}\right), \ v = \frac{5}{3}.$$

Игру с матрицей 2 x 2 можно решить графически, с помощью таких построений.





По оси абсцисс откладываем отрезок, длина которого равна единице. Левый конец отрезка (точка $x = 0$) отвечает стратегии A_1, правый – стратегии A_2. Промежуточные точки x отвечают смешанным стратегиям (x_1, x_2), где $x_1 = 1 - x$, $x_2 = x$. На концах этого отрезка перпендикулярно к оси абсцисс проводят прямые и откладывают выигрыш при соответствующих чистых стратегиях. Если игрок B использует стратегию B_1, то выигрыш при использовании чистых стратегий A_1 и A_2 составит соответственно a_{11} и a_{21}. Отложим эти точки и соединим их прямой B_1B_1. Если игрок A использует смешанную стратегию, то его выигрышу соответствует некоторая точка M, которая лежит на этой прямой. Аналогично можно построить прямую B_2B_2, соответствующую стратегии B_2 игрока B. Ломана B_1KB_2 – нижняя граница выигрышу, который получит игрок A. Точка K, в которой выигрыш максимален, определяет цену игры и ее решение. Для нахождения оптимальной стратегии игрока B используются формулы

$$y_1 = \frac{LB_2}{LB_2 + LB_1}, \quad y_2 = \frac{LB_1}{LB_2 + LB_1}.$$

Справедливость этих соотношений доказывается, если в формулы, выражающие y_1 и y_2, подставить вместо LB_2 и LB_1 их значения, имеем:

$$LB_2 = v - a_{22}\ ;\ LB_1 = a_{21} - v.$$

Здесь же можно рассмотреть задачу минимизации верхней границы выигрышу для игрока B, поменяв местами в решении игроков A и B.

Используя геометрическую интерпретацию, можно найти решение игр, заданных матрицей $2 \times n$. Каждой из n стратегий игрока B соответствует прямая. Построив эти прямые, находят нижнюю границу выигрыша. Точка K, лежащая на нижней границе и для которой величина выигрышу наибольшая, определяет цену игры и ее решение. При этом определяются активные стратегии игрока B (соответствующие им прямые пересекаются в точке K): из геометрических соображений можно найти значения y_j, соответствующие активным стратегиям игрока B.

Аналогично может быть решена игра с матрицей $m \times 2$, только при этом строят верхнюю границу выигрышу и на ней определяют минимум.

Геометрические построения целесообразно использовать для определения активных стратегий игроков. Потом решение находят по аналитическим формулам (1) – (3), или соответствующие значения **X, Y** и v находят из геометрических соображений.

Пример. Найти решение игры, заданной матрицей

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}.$$

Решение. Прямые на рисунке соответствуют стратегиям игрока B. Ломана B_3KB_4 соответствует нижней границе выигрыша. Оптимальные стратегии игрока B – третья и четвертая. По формулам (1) – (3) находим решение игры: $\mathbf{X} = (0{,}4;\ 0{,}6)$; $\mathbf{Y} = (0;\ 0;\ 0{,}6;\ 0{,}4)$; $v = 2{,}2$. Следовательно, игрок A использует стратегию A_1 с вероятностью 0,4, а стратегию A_2 – с вероятностью 0,6. При этом его выигрыш в среднем составит 2,2 единиц.

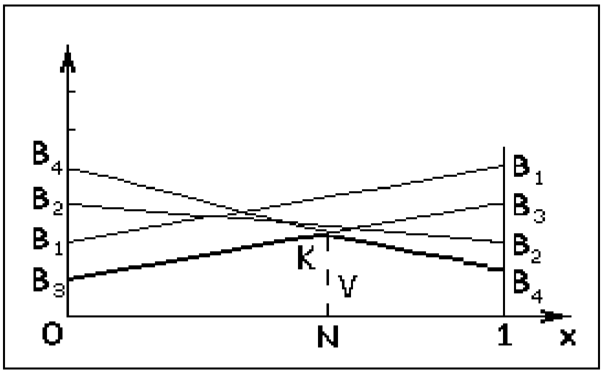


Рис. Графическое решение игры, матрица которой имеет размеры 2 х 4

Вопросы для самопроверки:

1) Какие задачи в экономике можно решать методами теории игр?
2) Привести основные понятия теории игр.
3) Как решить матричную игру в чистых стратегиях?

4) Как решить матричную игру в смешанных стратегиях?
5) Привести основную теорему теории игр.
6) Как решить матричную игру графическим методом?

Тема 15. Анализ и управление риском в экономике на базе концепции теории игр

15.1. Игры с природой
15.2. Критерии для принятия решений

15.1. Игры с природой

С целью уменьшения неблагоприятных последствий в каждом конкретном случае следует учитывать степень риска и имеющуюся информацию. Здесь лицо, принимающее решение (ЛПР), вступает в игровые отношения с некоторым абстрактным лицом, которое условно называют «природой». При этом имеются некоторые вероятностные характеристики состояния природы. Такую ситуацию принято называть играми с природой.

Любую хозяйственную деятельность человека можно рассматривать как игру с природой. В широком смысле под «природой» понимают совокупность неопределенных факторов, влияющих на эффективность принятых решений.

Задачей лица принимающего решение (ЛПР) является принятие наилучшего управленческого решения в каждой конкретной ситуации. Качество принимаемого решения зависит от информированности ЛПР о ситуации, в которой принимается решение. Безразличие природы к результату игры (выигрышу) и возможность получения ЛПР дополнительной информации о ее состоянии отличают игру с природой от обычной матричной игры.

Игры с природой представляют собой основную модель теории принятия решений в условиях частичной неопределенности.

Множество состояний природы обозначим через Π, отдельное состояние - $\Pi_j, \Pi_j \in \Pi$ $(j = \overline{1,n})$. Множество решений (стратегий) ЛПР обозначим через $A_i, A_i \in A$ $(i = \overline{1,m})$.

Во взаимоотношениях с природой ЛПР может использовать любые из стратегий $A_1, ..., A_m$ в зависимости от состояний Π_j природы. ЛПР отыскивает оптимальное поведение, которой и будет его оптимальной стратегией. При этом он может пользоваться как чистыми, так и смешанными стратегиями.

На основании платежной матрицы игры с природой

$$\begin{pmatrix} a_{11}\ a_{12} \dots a_{1n} \\ a_{21}\ a_{22} \dots a_{2n} \\ \dots\dots\dots\dots\dots \\ a_{m1}\ a_{m2} \dots a_{mn} \end{pmatrix},$$

где a_{ij} - выигрыш ЛПР, если он использует стратегию A_i при состоянии природы $П_j$.

Отличием игры с природой от обычной матричной игры состоит прежде всего в упрощении игры. Выявление дублирующих и доминирующих стратегий производится только для стратегий ЛПР.

Иногда при решении игры с природой используется матрица рисков. Элементы r_{ij} матрицы рисков равны разности между максимально возможным выигрышем и тем выигрышем, который статистик получит в тех же условиях $П_j$, применяя стратегию A_i, т.е. $r_{ij} = \beta_j - a_{ij}$, где $\beta_j = \max_i a_{ij}$.

15.2. Критерии для принятия решений

Оптимальную стратегию ЛПР можно определить, используя ряд критериев. Рассмотрим две группы критериев, использующих и не использующих априорные вероятности q_j состояний природы. К первой группе относятся критерии Байеса и Лапласа. В этом случае пользуются средним значением $\overline{a_i}$ выигрыша.

Так, при известном распределении вероятностей различных состояний $П_j$ природы используется критерий Байеса. Показателем в этом критерии служит либо величина среднего выигрыша, либо величина среднего риска.

Платежную матрицу $(a_{ij})_{m\times n}$ представляют в виде табл. 15.2.

Таблица 15.2

Стратегия ЛПР A_i	Состояние природы $П_j$				Средний выигрыш $\overline{a_i}$
	$П_1$	$П_2$	…	$П_n$	
A_1 … A_m	a_{11} … a_{m1}	a_{12} … a_{m2}	… … …	a_{1n} … a_{mn}	$\overline{a_1}$ … $\overline{a_m}$
q_j	q_1	q_2	…	q_n	

По критерию Байеса за оптимальную принимается та чистая стратегия A_i, при которой максимизируется средний выигрыш $\bar{a}_i$ ЛПР, т.е. обозначается $\bar{a} = \max\limits_i \bar{a}_i$, где $\bar{a}_i = \sum\limits_{j=1}^{n} a_{ij} q_j$.

Матрицу рисков представляют в виде таб. 15.2.

Таблица 15.2

Стратегия ЛПР A_i	Состояние природы $П_j$				Средний риск $\bar{r}_i$
	$П_1$	$П_2$	…	$П_n$	
A_1 … A_m	r_{11} … r_{m1}	r_{12} … r_{m2}	… … …	r_{1n} … r_{mn}	$\bar{r}_1$ … $\bar{r}_m$
q_j	q_1	q_2	…	q_n	

За оптимальную стратегию ЛПР принимается чистая стратегия A_i, при которой минимизируется средний риск, т.е. обеспечивается $\bar{r} = \min\limits_i \bar{r}_i$, где $\bar{r}_i = \sum\limits_{j=1}^{n} r_{ij} q_j$.

В случае, когда вероятности состояний природы правдоподобны, для их оценки используется принцип недостаточного основания Лапласа, согласно которому все состояния природы полагаются равновероятными, т.е. $q_1 = \ldots = q_n = \dfrac{1}{n}$. Оптимальной считается стратегия, обеспечивающая максимум среднего выигрыша.

Во вторую группу критериев выбора оптимальной стратегии ЛПР, используются при неизвестных априорных вероятностях состояний природы, относятся критерии Вальда, Сэвиджа и Гурвица.

Максимальный критерий Вальда совпадает с критерием выбора максимальной стратегии, позволяющей получать нижнюю чистую цену α в парной игре с нулевой суммой. При критерию Вальда за оптимальную принимается чистая стратегия, которая в наихудших условиях гарантирует максимальный выигрыш, т.е. $\alpha = \max\limits_i \min\limits_j \{a_{ij}\}$.

Критерий минимального риска Сэвиджа рекомендует выбирать в качестве оптимальной стратегии ту, при которой величина максимального риска минимизируется в наихудших условиях, т.е. обеспечивается $\min\limits_i \max\limits_j \{r_{ij}\}$.

Критерии Вальда и Сэвиджа ориентируют ЛПР на самые неблагоприятные состояния природы, т.е. эти критерии выражают пессимистическую оценку ситуации.

Критерий Гурвица является критерием пессимизма-оптимизма. За оптимальную принимается та стратегия, для которой выполняется соотношение

$$\max_i \left\{ \lambda \min_j \{a_{ij}\} + (1-\lambda) \max_j \{a_{ij}\} \right\}, \quad \text{де} \quad 0 \le \lambda \le 1.$$

При $\lambda = 0$ имеем критерий крайнего оптимизма, а при $\lambda = 1$ - критерий пессимизма Вальда. Если $0 < \lambda < 1$, то имеем среднее. При желании подстраховаться в данной ситуации λ принимают близким в единице. В общем случае число λ выбирают исходя из опыта или субъективных соображений.

Пример. Продолжим решение примера лекции 14.

Таблица 15.3

Матрица платежеспособного спроса

Объем произв (оптовые закупки)	Размер прибыли (a_{ij}) в зависимости от возможных колебаний спроса				$\alpha_i = \min \{a_{ij}\}$
	489 876,17	986 200,50	1488 973	1967 962,1	
S_1=986 200,50	49 310,03	197 240,1	197 240,1	197 240,1	49 310,03
S_2=1488 973	-60	148 897,3	297 794,6	297 794,6	-60
S_3=1967 962,1	-1140	98 398,11	196 796,24	393 592,42	-1140
$\beta_j = \max \{a_{ij}\}$	49 310,03	197 240,1	297 794,6	393 592,42	

Построим матрицу рисков.

Таблица 15.4

r_{ij}	489 876,17	986 200,50	1488 973	1967 962,1	$\max \{r_i\}$	$S_{опт}$
S_1	0	0	100554,5	196352,32	196352,32	
S_2	49370,03	48342,80	0	95797,82	95797,82	95797,82
S_3	50450,03	98841,99	100998,36	0	100998,36	

Смысл Сэвиджа критерия заключается в стремлении избежать большего риска при выборе решений. Следовательно, в соответствии с критерием, необходимо производить продукцию в объеме $S_2 = 1488973$ тыс. грн.

Пусть мы преследуем пессимистическую оценку (критерий Гурвица) и считаем, что $\lambda = 0{,}8$, тогда для каждой стратегии соответственно имеем:

$$y_1 = 0{,}8 \cdot 49310{,}03 + (1 - 0{,}8) \cdot 197240{,}1 = 78896{,}04 \text{ тис. грн},$$

$$y_2 = 0{,}8 \cdot (-60) + (1 - 0{,}8) \cdot 297794{,}6 = 59510{,}92 \text{ тис. грн},$$

$$y_3 = 0{,}8 \cdot (-1140) + (1 - 0{,}8) \cdot 393592{,}42 = 77806{,}48 \text{ тис. грн}.$$

По критерию Гурвица рациональным есть такой вариант объема производства:

$$Y = \max_i \{y_i\} = \max\{78896{,}04;\ 59510{,}92;\ 77806{,}48\} = 78896{,}04 \text{ тис. грн} \rightarrow S_1.$$

Вопросы для самопроверки:

1) Для каких задач в экономике применяются игры с природой?

2) Привести основные критерии для принятия решений в задачах игр с природой.

3) В чем суть критерия Байеса?

4) В чем суть критерия Вальда, Сэвиджа и Гурвица?

Тема 16. Динамическое программирование

16.1. Общие теоретические выкладки

16.2. Решение задачи динамического программирования в аналитической форме

16.3. Решение задачи о замене оборудования

16.1. Общие теоретические выкладки

Основным методом динамического программирования является разработанный американским математиком Р. Беллманом метод рекуррентных соотношений, в основе которого лежит принцип оптимальности: если управление процесса оптимально, то оно будет оптимальным и для процесса, остающегося после выполнения первого шага. Этот принцип позволяет установить соотношение между экстремальными значениями целевой функции в задачах с различной продолжительностью процесса в различными начальными состояниями [1, с. 125].

Математическая формулировка принципа оптимальности чаще приводится для задач с аддитивным критерием оптимальности (сепарабальная функция цели).

Пусть необходимо найти максимум функции:

$$Z = z_1(x_1) + z_2(x_2) + z_3(x_3)$$

при ограничениях

$$\begin{cases} u_1(x_1) + u_2(x_2) + u_3(x_3) \le a_0, \\ v_1(x_1) + v_2(x_2) + v_3(x_3) \le b_0 \end{cases}$$

и любым двухсторонними ограничениями на неизвестные.
Удобно интерпретировать x_j как объем производства определенного товара, $z_j(x_j)$ - прибыль от реализации этой продукции, $u_j(x_j)$, $v_j(x_j)$ - затраты двух видов ресурсов на производство, a_0, b_0 - имеющиеся запасы ресурсов.

Сведем эту многомерную задачу к последовательности одномерных задач, в каждой из которых учитывается только одна неизвестная. Обозначим через $f_1(a_1, b_1), f_2(a_2, b_2), f_3(a_3, b_3)$ - максимальную прибыль, полученную от выпуска только одного вида товара, только первых двух видов и всех трех наименований при заданных ресурсах в объемах $a_j \le a_0, b_j \le b_0$. При таких обозначениях оптимальное решение можно записать как $Z_{\max} = f_3(a_0, b_0)$, где a_0, b_0 - конкретные числовые значения объемов ресурсов, а a_j, b_j- еще неизвестные параметры.

Первый* (*начальный*) *этап. Пусть заданы все неизвестные, кроме первой, и необходимо найти оптимальный план выпуска продукции одного первого наименования при имеющихся остатках ресурсов a_1, b_1. Другими словами, необходимо найти $f_1(a_1, b_1) = \max\ \{z_1(x_1)\}$ при ограничениях $u_1(x_1) \le a_1$, $v_1(x_1) \le b_1$ и двухсторонних ограничениях на x_1.

Второй этап. Определим оптимальный план производства двух видов продукции при имеющихся остатках ресурсов a_2, b_2.

Необходимо найти $f_2(a_2, b_2) = \max\ \{z_1(x_1), z_2(x_2)\}$

при ограничениях

$$\begin{cases} u_1(x_1) + u_2(x_2) \le a_2, \\ v_1(x_1) + v_2(x_2) \le b_2 \end{cases}$$

и двухсторонних ограничениях на неизвестные.

Сведем эту двумерную задачу к одномерной. Пусть неизвестная x_2 – задана, тогда запасы ресурсов на производство оставшихся видов продукции сокращаются на $u_2(x_2)$, $v_2(x_2)$ и максимум прибыли (при заданной x_2) будет равен

$$z_2(x_2)+f_1(a_2-u_2(x_2),b_2-v_2(x_2)),$$

где $f_1(a_2-u_2(x_2),b_2-v_2(x_2))=f_1(a_1,b_1)$ найденная на предыдущем этапе наибольшая прибыль от выпуска первого вида товара при остатках ресурсов, которые сейчас зависят от заданного значения x_2.

Общая прибыль от выпуска двух видов товара оказалась выраженной через одну неизвестную x_2. Перебирая ее возможные значения, находим максимум прибыли при производстве двух видов продукции в такой форме:

$$f_2(a_2,b_2)=\max_{x_2}\left\{z_2(x_2)+f_1(a_2-u_2(x_2),b_2-v_2(x_2))\right\}.$$

Получили известное функциональное (реккурентное) уравнение Беллмана. Неизвестная x_2 – задается, далее добавляется оптимальное решение предыдущего этапа $f_1(a_1,b_1)$ для остатков ресурсов $a_1=a_2-u_2(x_2), b_1=b_2-v_2(x_2)$, $b_1 = b_2 - v(x_2)$. Разыскивается максимум полученной суммы по всем возможным x_2.

Третий этап. Повторяя рассуждения предыдущего этапа, получаем аналогичное функциональное уравнение:

$$f_3(a_3,b_3)=\max_{x_3}\left\{z_3(x_3)+f_2(a_3-u_3(x_3),b_3-v_3(x_3))\right\}.$$

Поскольку третий этап – заключительный, то можно не решать задачу с переменными запасами ресурсов, а сразу положить их конкретным числовых значениям заданных запасов ресурсов $a_3=a_0, b_3=b_0$. На заключительном этапе определяется последняя компонента оптимального плана x_3^0.

Обратным ходом процедуры определяем остальные компоненты оптимального плана. Поскольку уже известно значение x_3^0, находим остатки ресурсов $a_2=a_3-u(x_3^0), b_2=b_3-v(x_3^0)$ и из общего решения второго этапа находим значение второй компоненты оптимального плана x_2^0. Далее вычисляем $a_1=a_2-u(x_2^0), b_1=b_2-v(x_2^0)$ и из общего решения первого этапа находим первую компоненту x_1^0.

16.2. Решение задачи динамического программирования в аналитической форме

Найти максимум сепарабельной функции цели:

$$z = z(x_1) + z(x_2) = x_1 + (2x_2 - x_2^2)$$

при ограничениях

$$\begin{cases} 3x_1 + 2x_2 \le 6, \\ x_1 + 2x_2 \le 4, \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

Первый* (*начальный*) *этап. На 1–м этапе необходимо найти максимум одномерной функции

$$f_1(a_1, b_1) = \max z_1(x_1) = \max x_1$$

при ограничениях:

$$\begin{cases} 0 \le 3x_1 \le a_1, \\ 0 \le x_1 \le b_1. \end{cases}$$

Задача простая и ее легко решить аналитически для всех a_1, b_1. Переменная x_1 принимает максимальное значение на правой границе интервала варьирования в точке $x_1 = \min\left\{\frac{a_1}{3}, b_1\right\}$. Значение $f_1(a_1, b_1)$ равно этому же выражению.

Второй (заключительный) этап. Выписываем реккурентное соотношение:

$$\begin{aligned} f_2(a_2, b_2) &= \max_{x_2} \{z_2(x_2) + f_1(a_2 - 2x_2, b_2 - 2x_2)\} = \\ &= \max_{x_2} \{(2x_2 - x_2^2) + f_1(6 - 2x_2, 4 - 2x_2)\} \end{aligned}$$

Поскольку этот этап – заключительный, то подставляем сразу известные значения параметров $a_2 = 6, b_2 = 4$.

Находим $x_1 = \min\left\{\frac{a_1}{3}, b_1\right\} = \min\left\{\left(\frac{6 - 2x_2}{3}, 4 - 2x_2\right)\right\}$ для $0 \le x_2 \le 2$.

На интервале $0 \le x_2 \le 1{,}5$ справедливо неравенство $\frac{6-2x_2}{3} \le (4-2x_2)$, откуда $x_1 = \frac{6-2x_2}{3}$. На интервале $1{,}5 \le x_2 \le 2$ справедливо противоположное неравенство $\frac{6-2x_2}{3} \ge (4-2x_2)$, откуда $x_1 = (4-2x)$,

$f_1(6-2x_2\,, 4-2x_2) = \max x_1$:

$$f_1(6-2x_2, 4-2x_2) = \begin{cases} 2-\frac{2}{3}x_2\,, & (0 \le x_2 \le 1{,}5) \\ 4-2x_2\,, & (1{,}5 \le x_2 \le 2) \end{cases}.$$

С учетом выражения для f_1 получаем:

$$z_{\max} = f_2(6,4) = \max\begin{cases} (2x_2 - x_2^2) + (2-\frac{2}{3}x_2), & (0 \le x_2 \le 1{,}5) \\ (2x_2 - x_2^2) + (4-2x_2), & (1{,}5 \le x_2 \le 2) \end{cases} =$$

$$= \max\begin{cases} (2+\frac{4}{3}x_2 - x_2^2), & (0 \le x_2 \le 1{,}5) \\ (4-x_2^2), & (1{,}5 \le x_2 \le 2) \end{cases} =$$

$$= \max\begin{cases} 2{,}444, & (x_2 = 0{,}667) \\ 1{,}750, & (x_2 = 1{,}5) \end{cases} =$$

$$= 2{,}444 \quad (x_2 = 0{,}667)$$

При $x_2 \le 1{,}5$: $z_{\max} = 2 + \frac{4}{3}x_2 - x_2^2$. Максимум квадратичного трехчлена достигается при $x_2 = \frac{2}{3} = 0{,}667$ и равен $z_{\max} = \frac{22}{9} = 2{,}444$. При $x_2 \le 1{,}5$: $z = 4 - x_2^2$. Максимум этого выражения достигается при $x_2 = 1{,}5$ и равен $z = \frac{7}{4} = 1{,}75$. Большее из полученных значений равно 2,444, при этом $x_2^0 =$ $= {}^2\!/_3 = 0{,}667$. Обратным ходом получаем $x_1^0 = 2 - \frac{2}{3}x_2^0 = \frac{14}{9}$.

16.3. Решение задачи о замене оборудования

Одной из основных задач, которые решаются на предприятии с помощью оптимизационных методов есть задача о замене оборудования. Старение оборудования включает физический и моральный износ, в результате чего увеличиваются производственные затраты, затраты на ремонт и обслуживание, снижается производительность труда и ликвидная стоимость оборудования. Необходимо разработать оптимальную стратегию использования и своевременной замены устаревшего оборудования с целью снизить до минимума суммарные затраты на приобретение оборудования и его эксплуатацию, что эквивалентно, чтобы получить максимальную прибыль. Зависимости, описывающие рост эксплуатационных расходов, снижение производительности и ликвидной стоимости оборудования, могут быть заданы аналитически или в табличной форме.

В этой задаче последовательность этапов не может быть произвольной. Время является определяющим фактором, поэтому допустимы только две возможные последовательности этапов – от начала к концу (алгоритм «прямой прогонки»), или от конца к началу (алгоритм «обратной прогонки»). Решение при этом подразделяется на ряд этапов, каждый из которых соответствует очередному году эксплуатации.

Пусть оборудование эксплуатируется $n = 5$ лет, после чего продается. В начале каждого года необходимо принять решение – сохранить старое оборудование, или же заменить его новым (возможны также и другие решения, например, произвести ремонтные работы различной степени сложности). Стоимость нового оборудования p, ликвидная стоимость снижается с каждым проработанным годом вдвое и после t лет эксплуатации составляет $s(t) = p \cdot 2^{-t}$. Затраты на эксплуатацию в течение года зависят от возраста оборудования и увеличиваются с каждым проработанным годом на одну и ту же величину $u(t) = 0{,}15 p \cdot t$. Найдем оптимальную политику своевременной замены оборудования при таких условиях задачи.

Алгоритм прямой прогонки

В этом методе потребуется найти величину эксплуатационных затрат за весь срок службы оборудования, для чего придется находить сумму членов арифметической прогрессии:

$$v(t) = u(1) + u(2) + \ldots + u(t) = 0{,}15 p (1 + 2 + \ldots + t) = 0{,}075 pt (t+1).$$

Проверим выражение: за первый год работы нового оборудования эти расходы составят $0{,}75 p \cdot 1 \cdot 2 = 0{,}15 p$ – так и было задано по условию задачи. Таким образом, суммарные эксплуатационные расходы здесь являются квадратичной функцией возраста оборудования (t – количество проработанных лет).

Теперь введем обозначение и выведем основное реккурентное соотношение. Обозначим через f_k минимальные расходы на эксплуатацию и обновление оборудования в течение срока k лет, после чего оборудование продается. В этих обозначениях нашей целью является определение $f_n = f_5$.

В принципе оборудование можно эксплуатировать неограниченно долго, так что к концу k–го года возраст оборудования может быть любым в пределах $1 \le t \le k$. Если предположить, что известен этот возраст, то легко определяются общие затраты за все время работы оборудования $w(t) = p - s(t) + v(t)$, а затраты за предыдущие годы выражаются функцией f_{k-t}. Поскольку возраст t нам заранее неизвестен, перебираем все его возможные значения из интервала $1 \le t \le k$ и находим минимум следующего выражения:

$$f_k = \min \{w(t) + f_{k-t}\}.$$

Это есть искомое реккурентное соотношение (функциональное уравнение Беллмана). Для начала вычислительного процесса полагаем $f_0 = 0$.

Значения расходов (с точностью до множителя p) берем из табл. в зависимости от возраста $1 \le t \le 5$:

Таблица

t	$s(t)$	$v(t)$	$w(t)$
1	$1/2$ = 0,500	0,075·1·2 = 0,150	1 – 0,500 + 0,150 = 0,650
2	$1/4$ = 0,250	0,075·2·3 = 0,450	1 – 0,250 + 0,450 = 1,200
3	$1/8$ = 0,125	0,075·3·4 = 0,900	1 – 0,125 + 0,900 = 1,775
4	$1/16$ = 0,062	0,075·4·5 = 1,500	1 – 0,062 + 1,500 = 2,438
5	$1/32$ = 0,031	0,075·5·6 = 2,250	1 – 0,031 + 2,250 = 3,218

Последовательно вычисляем значения $f_k = \min \{w(t) + f_{k-t}\}$:

$$f_1 = \min \{w(1) + f_0\} = 0{,}650 + 0 = 0{,}650; \qquad t = 1.$$

$$f_2 = \min \begin{Bmatrix} w(1) + f_1 \\ w(2) + f_0 \end{Bmatrix} = \min \begin{Bmatrix} 0{,}650 + 0{,}650 \\ 1{,}200 + 0 \end{Bmatrix} = \min \begin{Bmatrix} 1{,}300 \\ 1{,}200 \end{Bmatrix} = 1{,}200; \quad t = 2.$$

$$f_3 = \min \begin{Bmatrix} w(1) + f_2 \\ w(2) + f_1 \\ w(3) + f_0 \end{Bmatrix} = \min \begin{Bmatrix} 0{,}650 + 1{,}200 \\ 1{,}200 + 0{,}650 \\ 1{,}775 + 0 \end{Bmatrix} = \min \begin{Bmatrix} 1{,}850 \\ 1{,}850 \\ 1{,}775 \end{Bmatrix} = 1{,}775; \quad t = 3.$$

$$f_5 = \min\begin{Bmatrix} w(1) + f_4 \\ w(2) + f_3 \\ w(3) + f_2 \\ w(4) + f_1 \\ w(5) + f_0 \end{Bmatrix} = \min\begin{Bmatrix} 0,650 + 2,400 \\ 1,200 + 1,775 \\ 1,775 + 1,200 \\ 2,438 + 0,650 \\ 3,218 + 0 \end{Bmatrix} = \min\begin{Bmatrix} 3,050 \\ 2,975 \\ 2,975 \\ 3,050 \\ 3,218 \end{Bmatrix} = 2,975; \quad \begin{matrix} t = 2, \\ t = 3. \end{matrix}$$

Задача решена. Оптимальная стратегия при таком соотношении расходов заключается в замене оборудования после каждых двух лет работы. Поскольку нормативный срок n = 5 лет оказался нечетным, предложено два равноценных решения – первый раз оборудование работает три года, новое оборудование работает два года, или наоборот, два года и три года после обновления. Минимальные расходы на эксплуатацию и обновление при такой стратегии составляют $2,975\,p$.

Алгоритм обратной прогонки

Обозначим через $\varphi_k(\tau)$ – минимальные расходы за последние k лет при условии, что в начале этого периода имеется оборудование возраста τ. Тогда $\varphi_1(\tau)$ – минимальные расходы за последний год, где $1 \le \tau \le n-1$; $\varphi_2(\tau)$ – минимальные расходы за последние два года, где $1 \le \tau \le n-2$; и т.д. до $\varphi_n(0)$ – искомые минимальные расходы за все n лет (в самом начале оборудование новое). На каждом этапе надо рассматривать две альтернативы – сохранить старое оборудование, или заменить его.

Пусть $k = 1$. Если оборудование возраста τ сохранить, годовые эксплуатационные расходы будут равны $u(\tau+1)$. Оборудование после этого постареет на год и его ликвидная стоимость будет равна $s(\tau+1)$. Последний год работы несколько отличается от остальных, поскольку по условиям задачи после $n = 5$ лет оборудование продается по остаточной стоимости независимо от его возраста, поэтому при сохранении оборудования в последний год работы общие расходы составят $u(\tau+1) - s(\tau+1)$. Если же в последний год оборудование заменить, общие расходы будут равны $p - s(\tau) + u(1) - s(1)$ – стоимость нового оборудования за вычетом ликвидной стоимости старого, плюс эксплуатационные расходы за 1 год работы нового оборудования, наконец, возврат денег от продажи только что купленного оборудования. Окончательно получаем:

$$\varphi_1(\tau) = \min \begin{cases} u(\tau+1) - s(\tau+1) & \text{– сохранить,} \\ p - s(\tau) + u(1) - s(1) & \text{– заменить.} \end{cases}$$

Все вычисления сведены в табл., где для каждого значения возраста оборудования τ вычислены ликвидная стоимость оборудования $s(\tau)$, эксплуатационные затраты за предыдущий год работы $u(\tau)$, общие затраты, если это оборудование сохранить, общие затраты, если оборудование заменить, и минимальные затраты в этих двух альтернативных вариантах.

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1)-s(\tau+1)$	$p+u(1)-s(1)-s(\tau)$	$\varphi_1(\tau)$
1	0,500	0,15	0,30–0,250=0,050	0,65–0,500=0,150	0,050 (сохранить)
2	0,250	0,30	0,45–0,125=0,325	0,65–0,250=0,400	0,325 (сохранить)
3	0,125	0,45	0,60–0,062=0,538	0,65–0,125=0,525	0,525 (заменить)
4	0,062	0,60	0,75–0,031=0,719	0,65–0,062=0,588	0,588 (заменить)
5	0,031	0,75			

К последнему 5-му году работы возраст оборудования может быть любым от 1-го до 4-х лет. В табл. Присутствует 5-я полупустая строка для вычисления $u(\tau+1)$ и $s(\tau+1)$ при $\tau = 4$. В последнем столбце табл. Вычислены минимальные затраты при выборе найлучшего варианта замены оборудования, откуда следует, что даже вначале последнего года работы оборудования целесообразно заменить, если его возраст больше 2-х лет.

Пусть теперь $k = 2$, т.е. рассматриваются последние два года работы. Этот этап самый важный для понимания, остальные этапы будут однотипными. Итак, как обычно, в начале этого периода оборудование можно сохранить или заменить. Если оборудование возраста τ сохранить, то эксплуатационные расходы в 1–й год работы будут равны $u(\tau+1)$, далее оборудование постареет на год и дальнейшие расходы в последний год (при оптимальной политике замены изношенного оборудования) составят $\varphi_1(\tau+1)$. Если же оборудование в начале периода заменить, то общие расходы будут равны $p - s(\tau) + u(1) + \varphi_1(1)$ – стоимость нового оборудования за вычетом ликвидной стоимости старого, плюс эксплуатационные расходы за 1 год работы нового оборудования, наконец, минимальные расходы за следующий год. Окончательно получаем:

$$\varphi_2(\tau) = \min \begin{cases} u(\tau+1) + \varphi_1(\tau+1) & \text{– сохранить,} \\ p - s(\tau) + u(1) + \varphi_1(1) & \text{– заменить.} \end{cases}$$

С точностью до множителя p эти вычисления сведены в таблице.

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1)+\varphi_1(\tau+1)$	$p+u(1)+\varphi_1(1)-s(\tau)$	$\varphi_2(\tau)$

1	0,500	0,15	0,30+0,325=0,625	1,20–0,500=0,700	0,625 (сохранить)
2	0,250	0,30	0,45+0,525=0,975	1,20–0,250=0,950	0,950 (заменить)
3	0,125	0,45	0,60+0,588=1,188	1,20–0,125=1,075	1,075 (заменить)
4	0,062	0,60			

Аналогично при $k=3$ получаем рекуррентное соотношение:

$$\varphi_3(\tau)=\min\begin{cases} u(\tau+1)+\varphi_2(\tau+1) & -\text{сохранить},\\ p-s(\tau)+u(1)+\varphi_2(1) & -\text{заменить}.\end{cases}$$

В таблице вычислены значения этих затрат (без множителя p).

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1)-\varphi_2(\tau+1)$	$(1+u(1)+\varphi_2(1))-s(\tau)$	$\varphi_3(\tau)$
1	0,500	0,15	0,30+0,950=1,250	1,775–0,500=1,275	1,250 (сохранить)
2	0,250	0,30	0,45+1,075=1,525	1,775–0,250=1,525	1,525 (сохран/замена)
3	0,125	0,45			

При $k=4$ имеем рекуррентное соотношение:

$$\varphi_4(\tau)=\min\begin{cases} u(\tau+1)+\varphi_3(\tau+1) & -\text{сохранить},\\ p-s(\tau)+u(1)+\varphi_3(1) & -\text{заменить}.\end{cases}$$

Вычисляем $\varphi_4(\tau)$- минимальные затраты за четыре последних года работы (табл.)

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1)+\varphi_3(\tau+1)$	$(1+u(1)+\varphi_3(1))-s(\tau)$	$\varphi_4(\tau)$
1	0,500	0,15	0,30+1,525=1,825	2,400–0,500=1,900	1,825 (сохранить)
2	0,250	0,30			

Вычисляем минимальные затраты за 5 лет работы – это только один вариант:

$$\varphi_5(0)=p\cdot(1+u(1)+\varphi_4(1))=(1+0{,}15+1.825)p=2{,}975p.$$

Оборудование следует заменять после каждых двух лет работы.

Минимальные затраты на эксплуатацию и замену оборудования равны $\varphi_5(0)=2{,}975p$.

Алгоритмы прямой и обратной прогонки настолько отличаются друг от друга, что свидетельствует, что методика динамического программирования еще далека от формального завершения, в ней многое зависит от искусства, от умения увидеть простейший путь решения.

Вопросы для самопроверки:

1) Какие общие теоретические выкладки методов динамического программирования?

2) Привести этапы реализации методов динамического программирования?

3) Привести сущность этапов решения задачи о замене оборудования.

4) В чем идея алгоритма прямой прогонки?

5) В чем идея алгоритма обратной прогонки?

Содержание

Раздел 2. Эконометрика

Тема 1. Особенности эконометрических моделей и принципы их построения

1.1. Особенности эконометрических моделей. Роли и место эконометрических моделей в анализе социально-экономических систем

1.2. Формирование совокупности наблюдений. Понятие однородности наблюдений. Точность исходных данных

1.3. Основные этапы построения эконометрических моделей. Особенности обоснования формы эконометрической модели

1.1. Особенности эконометрических моделей. Роли и место эконометрических моделей в анализе социально-экономических систем.

В экономико-математическом энциклопедическом словаре эконометрика определена как научная дисциплина, позволяющая на базе положений экономической теории и результатов экономических измерений придавать конкретные количественные выражения общим (количественным) закономерностям, обусловленным экономической теорией [1, с. 597]. Эконометрика основывается на математической статистике, прежде всего, многомерного статистического анализа, экономической теории и экономической статистики. Так, математическая статистика обеспечивает прикладным математическим инструментарием, экономическая теория нацеливает на выявление объективно существующих экономических законов между экономическими показателями, указывает направления формализации, методы спецификации и идентификации соответствующих эконометрических моделей, экономическая статистика обеспечивает информацией эконометрическое моделирование.

Предметом эконометрики являются экономические и социально-экономические приложения, а именно модельное описание конкретных количественных взаимосвязей, существующих между анализируемыми показателями [1, с. 597].

Результатом применения инструментов эконометрики есть построение эконометрической модели. Эконометрическая модель содержит набор регрессионных уравнений, описывающих стохастические связи между исследуемыми экономическими показателями, а также отдельные тождества, характеризующие соотношения между экономическими показателями.

Считается, что наиболее распространенный математический вид исследуемых связей – линейная (относительно анализируемых переменных) и аддитивные формы. При этом возможны ситуации, когда одни и те же показатели в одних уравнениях модели есть объясняемыми, а в других – объясняющие (такие модели принято называть системами одновременных уравнений) [34, с. 619].

Эконометрическое моделирование нацелено на получение двух типов конечных прикладных результатов:

- описание различных состояний социально-экономических систем (СЭС), имитация разных возможных сценариев их развития;
- прогнозирование экономических показателей, характеризующих состояние и развитие СЭС.

Эконометрические модели используются для анализа социально-экономических систем на всех уровнях их управления, а именно на макро уровне – страны мира, мировой экономике, на мезо уровне – регионы внутри страны, административные районы и на микро уровне – предприятия, различные структуры бизнеса, домашнее хозяйство. С помощью эконометрических моделей проводится анализ всех видов деятельности субъектов хозяйствования.

В эконометрическом моделировании принято различать основные типы переменных: эндогенные, экзогенные, лаговые. Под экзогенными переменные понимаются признаки или показатели, значения которых задаются извне модели и они являются управляемыми или планируемыми. Эндогенные переменные формируются в процессе функционирования социально-экономической системы и зависят от экзогенных переменных или от их взаимодействия. Лаговые переменные есть эндогенные переменные, значения которых измеренные в прошлые моменты времени и являются уже заданными.

1.2. Формирование совокупности наблюдений. Понятие однородности наблюдений. Точность исходных данных.

Основой эконометрических моделей являются статистические данные, которые формирует экономическая статистика. Статистическая база для эконометрической модели может состоять из временных и пространственных (структурных) рядов данных. *Временным рядом* называется ряд значений статистического показателя, упорядоченного по времени. Различают краткосрочные (до 1 года), среднесрочные (1-3 года) и долгосрочные (свыше 5 лет) временные ряды.

В эконометрическом исследовании используют понятие относительной однородности. При формировании совокупности наблюдений, следует соблюдать сравнимость данных в пространстве и во времен. Это обозначает, что исходные данные должны иметь:

1) одинаковую степень агрегирования;

2) однородную структуру единиц совокупности;

3) одни и те же методы расчета показателей во времени;

4) одинаковую периодичность учета частных переменных;

5) сравнительные цены и одинаковые другие внешние экономические условия.

Критерии математической статистики позволяют количественно проверить однородность. Это критерий Фишера (рассматривался в теории вероятностей и математической статистике) и критерий Бартлетта.

Критерий Бартлетта есть наиболее общим критерием в решении проблем однородности совокупности наблюдений:

$K_Б = \frac{A}{B}$, $A = k\,n\ln\sigma^2_{общ} - n\sum_i \ln\sigma_i^2$; $B = 1 + \frac{k+1}{3nk}$, где k – количество групп, n – общее количество наблюдений; $\sigma^2_{общ}$ - общая дисперсия, σ_i^2 - дисперсия i-й группы наблюдений. Если $\frac{A}{B}$ не превышает χ^2 при выбранной уровней значимости α и $k-1$ степеней свободы , то можно утверждать, что дисперсии взяты с однородной совокупности наблюдений.

Точность измерения – это его адекватность. Универсальные критерии точности отсутствуют. Критерий точности каждого вида измерения определяется в соответствии с целями этого измерения. Погрешности измерения не сводятся в арифметическим погрешностям.

1.3. Основные этапы построения эконометрических моделей. Особенности обоснования формы эконометрической модели.

При построении эконометрической модели следует придерживаться общих методологических принципов экономико-математического моделирования. В соответствии с эти процедура построения эконометрической модели должна содержать шесть основных этапов:

1) постановочный этап, предполагающий определение конечных целей моделирования, набора признаков или показателей;

2) априорный этап, где проводится теоретико-логический анализ социально-экономической системы и ее частей, генезис исходных данных;

3) параметризация, когда выполняется собственно моделирование, то есть выбор общего вида модели, в том числе состава и формы входящих в нее связей;
4) информационный этап, предполагающий сбор необходимой статистической информации;
5) идентификация модели, согласно который выполняется статистический анализ модели, прежде всего, статистическое оценивание неизвестных параметров модели;
6) верификация модели предусматривает сопоставление реальных и модельных данных, проверку адекватности модели, оценку точности модельных данных.

Цель построения эконометрической модели – количественное описания взаимосвязей между переменными, поэтому она прежде всего строится с помощью методов регрессии и корреляции. При этом простая регрессия имеет вида: $\hat{y}_x = f(x)$, множественная: $\hat{y} = f(x_1, x_2, ..., x_n)$.

Формирование спецификации модели – это формулировка вида модели, исходя из соответствующей экономической теории связи между переменными. Различают объясняющие переменные – факторы. Факторы влияют на результат, поэтому выделяют результативный признак. Из круга факторов, влияющих на результативный признак, необходимо выделить наиболее существенно влияющие факторы. Предположим, что выдвигается гипотеза: величина спроса y на товар находится в обратной зависимости от цены x, т.е. $\hat{y}_x = a - bx$, при этом необходимо знать, какие остальные факторы предполагаются неизменными, возможно, в дальнейшем их придется учесть в модели и от простой регрессии перейти к множественной.

Практически в каждом отдельном случае величина результативного признака складывается из двух слагаемых: $y_j = \hat{y}_j + \varepsilon_j$, где y_j - фактическое значение результативного признака; $\hat{y}_j$ - теоретическое значение результата, найденное из уравнения регрессии; ε_j - случайная величина, отклонение реального значения от теоретического. Если правильно выбранная спецификация модели, то величина случайных ошибок меньше.

В парной регрессии выбор математической функции $\hat{y}_x = f(x)$ может быть осуществлен тремя методами: графическим, аналитическим, экспериментальным.

Графический метод основан на поле корреляции. Аналитический метод основан на изучении материальной природы связи исследуемых признаков. При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т.е. путем сравнения величины остаточной дисперсии $D_{\text{ост}} = \frac{1}{n}\sum(y - \hat{y}_x)^2$, рассчитанной при разных моделях. Чем меньше величина остаточной дисперсии, тем в меньшей мере наблюдается влияние не учитываемых в уравнении регрессии факторов и лучше уравнение регрессии подходит к исходным данным. При обработке статистических данных на компьютере перебираются разные математические функции в автоматическом режиме и из них выбирается та, для которой остаточная дисперсия является наименьшей.

Методы верификации основаны на процедурах статистической проверки гипотез и на статистическом анализе характеристик точности различных статистических приемов оценивания параметров. Наиболее распространенным и эффективным подходом к верификации эконометрической модели признан принцип так называемых ретроспективных расчетов, которые зависят от конечных целей моделирования.

Обычно эконометрика включает четыре раздела: изложение методов построения классической регрессионной модели и применение классического метода наименьших квадратом; построение обобщенной линейной регрессии и применение обобщенного метода наименьших квадратов; статистический анализ временных рядов; анализ систем одновременных уравнений.

Вопросы для самопроверки:

1) Как определяется эконометрика? На какой научной основе она строится?
2) В чем отличие эконометрической модели от других моделей?
3) Какие основные цели разработки эконометрических моделей?
4) Какие типы переменных различают в эконометрическом моделировании?
5) В чем суть понятия однородности наблюдений? Какие требования предъявляются к исходным данным при соблюдении однородности?
6) Какие основные этапы построения эконометрических моделей?
7) Раскрыть суть процесса формирования спецификации эконометрической модели.
8) Из каких разделов состоит эконометрика и, соответственно, какие типы эконометрических моделей различают?

Тема 2. Парная регрессия и корреляция в эконометрических исследованиях

2.1. Общие понятия регрессионного анализа. Типы связей

2.2. Линейная регрессия и корреляция: содержание и оценка параметров. Оценивание параметров линейной модели парной регрессии с помощью метода наименьших квадратов (МНК)

2.3. Примеры определения параметров МНК

2.4. Нелинейная регрессия

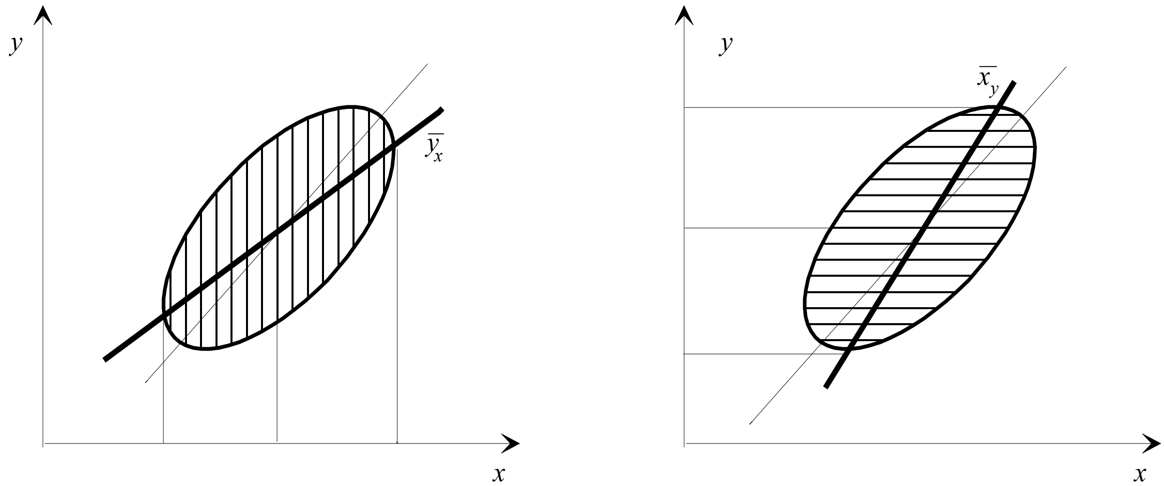
2.1. Общие понятия регрессионного анализа. Типы связей

Корреляционно-регрессионный анализ является одним из основных методов построения эконометрической модели. Цель корреляционно-регрессионного анализа заключается в установлении факта наличия или отсутствия зависимостей между несколькими показателями и описании этих связей достаточно простыми выражениями. Среди всех заданных показателей (признаков, переменных) один показатель считается результативным признаком y (откликом) и на этот показатель оказывают влияние остальные объясняющие переменные x_j факторы.

По определению *функциональных связей*, каждому значению аргумента (набору значений аргументов) соответствует единственное значение результативного признака. В природе часто встречается другой тип зависимостей, когда одному и тому же набору аргументов каждый раз соответствует несколько различные значения отклика (имеется случайный разброс измеряемого показателя). В таких *стохастических* (*вероятностных*) зависимостях каждому значению аргумента соответствует свой ряд распределения результативного признака. При изменении аргумента X меняются все характеристики распределения Y – меняется его центр группировки, разброс, вид распределения. В корреляционно-регрессионном анализе мы будем следить за изменением только одной характеристики распределения зависимой переменной – центра группировки Y при каждом значении X (то есть за изменением условного математического ожидания $M(y \mid x)$ при изменении аргумента x): $M(y \mid x)=f(x)$. Такой частный вид стохастической зависимости называется *корреляционной зависимостью*, ее график - *линией регрессии*, уравнение - *уравнением регрессии*.

Для корреляционных, так и для функциональных зависимостей, имеет место однозначное соответствие между значениями аргумента x и средними значениями результативного признака $\bar{y}_x$ (выборочная оценка условного математического ожидания). Однако между этими видами зависимостей остается принципиальное отличие – корреляционные зависимости необратные относительно замены направления причинно-следственных связей.

Рис. 2.1. демонстрирует различия между так называемыми сопряженными линиями регрессии $\bar{y}_x = f_1(x)$ и $\bar{x}_y = f_2(y)$.



а *б*

Рис. 2.1. Сопряженные линии регрессии: *а*) x - аргумент, y - функция; *б*) y - аргумент, x - функция.

Данный пример соответствует наиболее распространенному случаю совместного нормального распределения двух случайных величин (X,Y). При уровне доверия P=0,95 практически все точки (95%) попадают в эллиптическую область, так как для двумерного нормального закона облако рассеивания точек (X,Y) имеет форму вытянутого эллипса.

Для функциональной зависимости не имеет значения, относительно какой переменной решено уравнение – график от этого не изменится. В регрессионном анализе исследователь должен сам назначить, какой признак является результативным, а какой факторным.

Обе сопряженные линии регрессии не совпадают с главной осью эллипса рассеивания данных – это последствие выбора только одной из переменных как случайной, которой согласно гипотезам Гаусса-Маркова приписывают все погрешности – как случайные, так и не случайные. В этих теоретических зависимостях предполагается, что одна переменная определяется другую. Однако бывает, что обе переменные (x, y) являются разными последствиями одной общей причины (например, вследствие временного тренда), что и обуславливает связь между этими переменными. В таких задачах обе переменные равноправные и случайные с совместным законом распределения системы случайных величин. Для описания таких зависимостей разработан специальный математический инструмент – так называемая диагональная регрессия Фриша. График диагональной регрессии совпадает с главной осью рассеивания данных. Следует отметить, что диагональная регрессия не есть регрессией по определению, этот ти связи отличается от корреляционного, поскольку точки диагонали эллипса не являются средними значениями одной из переменных при заданном значении другой переменной.

Различают линейные и нелинейные регрессии. Линейная регрессия нашла широкое применение в эконометрике в виде четкой экономической интерпретации ее параметров. Реальные значения зависимой переменной не всегда совпадают с ее условными математическими ожиданиями и могут быть различными при одном и том же значении объясняющей переменной, поэтому зависимость должна быть дополнена слагаемым ε, что есть случайной величиной и указывает на стохастическую суть зависимости. Присутствие в модели случайной величины ε обусловлено тремя причинами: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

Линейная парная регрессия сводится к нахождению уравнения вида: $\hat{y}_x = a + b \cdot x$ или $y = a + b \cdot x + \varepsilon$. Зависимость нескольких переменных, выражаемая функцией $M(Y \mid x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n)$, называют множественной регрессией.

2.2. Линейная регрессия и корреляция: содержание и оценка параметров. Оценивание параметров линейной модели парной регрессии с помощью метода наименьших квадратов (МНК).

Уравнение вида $\hat{y}_x = a + b \cdot x$ позволяет по заданным значениям фактора x вичислить теоретические значения результативного признака. На графике теоретические значения представляют линию регрессии (рис 2.2.) [31, с. 41].

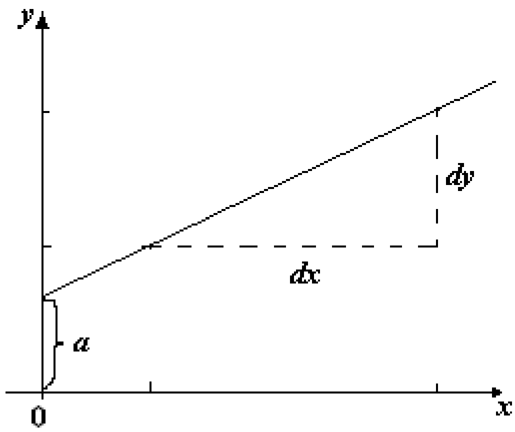


Рис. 2.2. Графическая оценка параметров линейной регрессии

Построение линейной регрессии сводится к оценке ее параметров a и b. Известные специалисты по проблемам корреляционно-регрессионного анализа рекомендуют постоянную регрессии a рассматривать как коэффициент при фиктивной переменной, принимающей во всех наблюдениях значение 1. Постоянная a определяет точку пересечения прямой регрессии с осью ординат. Поскольку в соответствии с общим истолкованием уравнения регрессии a является средним значением y в точке $x=0$, отсюда следует, что экономическая интерпретация a часто очень затруднительна или вообще невозможна [30, с. 56 – 57]. Постоянная a выполняет в уравнении регрессии функцию выравнивания и, благодаря ей, функция регрессии неошибочна. Поскольку уравнение регрессии интерпретируемо только в области скопления точек, а следовательно, только между наименьшим и наибольшим наблюдаемыми значениями переменной x, то постоянную a не обязательно интерпретировать. В экономических исследованиях у параметра a интерпретируется только знак: если a > 0, то относительное изменение результата происходит медленнее, чем изменение фактора.

Параметр b называется коэффициентом регрессии. Он характеризует наклон прямой к оси абсцисс или $b = tg\alpha$, где α- угол, который прямая регрессия образует с осью абсцисс. Коэффициент регрессии является мерой зависимости переменной y от переменной x или мерой влияния, оказываемого изменением переменной x на переменную y. Коэффициент b указывает среднюю величину изменения переменной y при изменении объясняющей переменной x на одну единицу, при этом знак b указывает направление этого изменения. Если $b > 0$ имеем положительную линейную регрессию, демонстрирующую поступательный характер изменения зависимой переменной при увеличении значений объясняющей переменной x. Если $b < 0$ имеем отрицательную линейную регрессию, при которой с увеличением значений x значения переменной y убывают. Постоянная a имеет ту же размерность, что и переменная y, а размерность коэффициента регрессии b представляет собой отношение размерности зависимой переменной к размерности объясняющей переменной.

Классический подход к оцениванию параметров линейной регрессии основан на ***методе наименьших квадратов*** (МНК), позволяющем получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}_x$ минимальна:

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2 \rightarrow \min, \varepsilon_i = y_i - \hat{y}_{x_i}, \quad \sum_{i=1}^{n}\varepsilon_i^2 \rightarrow \min .$$

Обозначим $S = \sum \sum \left(y_i - \hat{y}_{x_i}\right)^2 = \sum\left(y - a - b \cdot x\right)^2$. Чтобы найти минимум данной функции, надо вычислить частные производные по каждому из параметров a и b и приравнять их к нулю.

$$\frac{dS}{da} = -2\sum y + 2 \cdot n \cdot a + 2 \cdot b\sum x = 0;$$

$$\frac{dS}{db} = -2\sum y \cdot x + 2 \cdot a\sum x + 2 \cdot b\sum x^2 = 0 .$$

Получим следующую систему нормальных уравнений для оценки параметров a и b:

$$\begin{cases} n \cdot a + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum y \cdot x. \end{cases}$$

Методом исключения находим: $b=\dfrac{n\sum\limits_{i=1}^{n}x_i y_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2}$, $a=\dfrac{\sum\limits_{i=1}^{n}y_i-b\sum\limits_{i=1}^{n}x_i}{n}$.

Или $a=\bar{y}-b\cdot\bar{x}$, $b=\dfrac{\overline{yx}-\bar{y}\bar{x}}{\overline{x^2}-\bar{x}^2}$, где $\bar{y}=\dfrac{\sum\limits_{i=1}^{n}y_i}{n}$ и $\bar{x}=\dfrac{\sum\limits_{i=1}^{n}x_i}{n}$, $\overline{yx}=\dfrac{\sum\limits_{i=1}^{n}y_i x_i}{n}$

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции: $r_{xy}=\dfrac{n\sum\limits_{i=1}^{n}x_i y_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{\sqrt{\left(n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2\right)\left(n\sum\limits_{i=1}^{n}y_i^2-\left(\sum\limits_{i=1}^{n}y_i\right)^2\right)}}$.

Существует связь $r_{xy}=b\dfrac{\sigma_x}{\sigma_y}=\dfrac{\overline{yx}-\bar{y}\bar{x}}{\sigma_x\sigma_y}$ $(-1\le r_{xy}\le 1)$.

Для оценки качества подбора линейной функции используют квадрат линейного коэффициента корреляции r_{xy}^2, который называется коэффициентом детерминации.

2.3. Примеры определения параметров МНК

Изучается зависимость себестоимости единицы изделия (y, тыс. грн) от величины выпуска продукции (x, тыс. шт.) по группам предприятий за отчетный период. Экономист обследовал n=5 предприятий и получил результаты в табл. 2.1.

1) Полагая, что между переменными x, y имеет место линейная зависимость, определить выборочное уравнение линейной регрессии.

Таблица 2.1

Номер предприятия	x	y	x^2	xy
1	2	1,9	4	3,8
2	3	1,7	9	5,1
3	4	1,8	16	7,2
4	5	1,6	25	8
5	6	1,4	36	8,4
Сумма	20	8,4	90	32,5

$$b=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}=\frac{5\cdot 32{,}5-20\cdot 8{,}4}{5\cdot 90-20^2}=-0{,}11;$$

$$a=\frac{\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i}{n}=\frac{8{,}4-(-0{,}11)\cdot 20}{5}=2{,}12;$$

$$y=a+bx=2{,}12+(-0{,}11)x.$$

2) Найти остатки ε, коэффициент корреляции Пирсона и коэффициент детерминации.

Номер	x	y	y^2	$\hat{y}=2{,}12+(-0{,}11)x$	$\varepsilon=y-\hat{y}$
1	2	1,9	3,61	1,90	0,00
2	3	1,7	2,89	1,79	-0,09
3	4	1,8	3,24	1,68	0,12
4	5	1,6	2,56	1,57	0,03
5	6	1,4	1,96	1,46	-0,06
Сумма	20	8,4	14,26		

$$r=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{\left(n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right)\left(n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2\right)}}=\frac{5\cdot 32{,}5-20\cdot 8{,}4}{\sqrt{(5\cdot 90-20^2)(5\cdot 14{,}26-8{,}4^2)}}=-0{,}904$$

.

Значение коэффициента корреляции близко к -1, что свидетельствует об очень сильной отрицательной связи (с ростом x значения y убывают). Знаки $b=-0{,}11$ и $r=-0{,}904$ совпадают. Коэффициент детерминации $r^2=(-0{,}904)^2=0{,}817$, то есть 81,7% общей вариации себестоимости y зависит от выпуска продукции x. Построенная модель не объясняет 18,3% вариации себестоимости. Эта часть вариации объясняется факторами, не включенными в модель.

3) Выполнить прогноз на основе построенной линейной модели.

Найдем прогнозное значение себестоимости y при выпуске продукции $x=5{,}5$ тыс. шт.

$y=2{,}12+(-0{,}11)x$;

$y(5{,}5) = 2{,}12 - 0{,}11 \cdot 5{,}5 = 1{,}515$ тыс. грн.

2.4. Нелинейная регрессия

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, нелинейные по объясняющим переменным:

- полиномы разных степеней, например, $y = a + b_1 x + b_2 x^2 + b_3 x^3 + \varepsilon$;

- равносторонняя гипербола $y = a + \dfrac{b}{x} + \varepsilon$.

Регрессии, нелинейные по оцениваемым параметрам:

- степенная $y = a \cdot x^b \cdot \varepsilon$;

- показательная $y = a \cdot b^x \cdot \varepsilon$;

- экспоненциальная $y = e^{a + b \cdot x} \cdot \varepsilon$.

Известно, что успешного использования МНК желательно, чтобы модель была линейной относительно параметров. Для двухпараметрических моделей, которые линейно зависят от параметров или могут быть приведены к такой форме преобразованиями, существует графический способ проверки их пригодности для описания данных (проверки адекватности модели).

Пусть $Y = F(x, y)$ и $X = \Phi(x, y)$ - такие функциональные преобразования, после которых форма связи формально приобретает линейный вид $Y = a + bX$. Следует отметить, что графиком линейной зависимости есть прямая, которая выделяется среди множества других кривых. Отсюда следует вывод, что если эмпирические точки в преобразованных координатах не группируются вокруг некоторой прямой, той принятая форма связи должна быть отклонена. С применением современных программных продуктов графики легко построить и преобразовать, поэтому с применением компьютера такой способ идентификации является чрезвычайно эффективным.

Чаще всего применяется или логарифмирование, или переход к обратным величинам. На рис. 2. 3 приведены дополнительные сведения о двухпараметрических линейных зависимостях, которые могут быть сведены к линейным отмеченными функциональными преобразованиями.

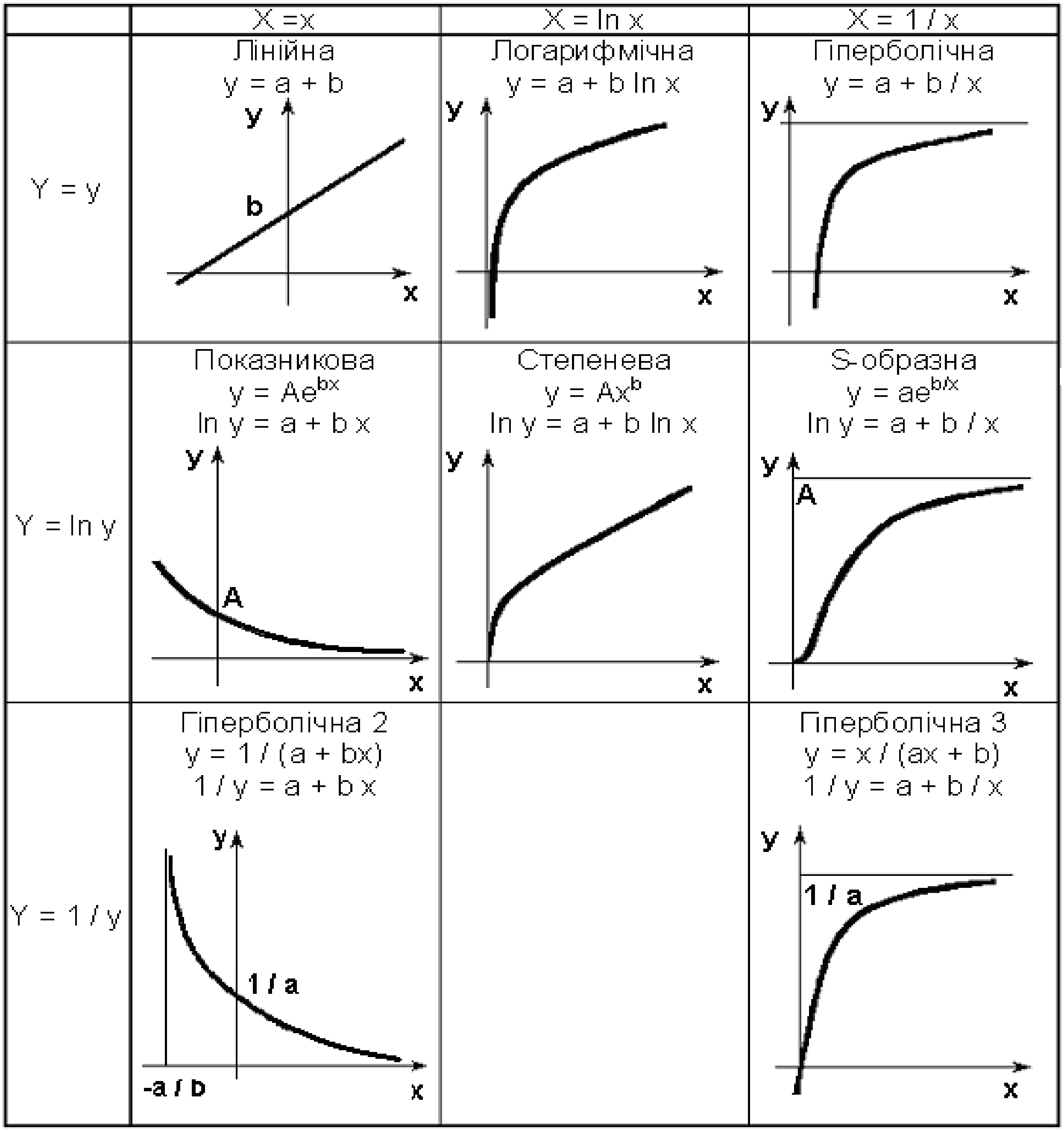


Рис. 2.3. Двухпараметрические зависимости $Y = a + bX$

С рис. 3 видно, что среди двухпараметрических зависимостей отсутствуют формы связи с экстремумами. Это является определенным ограничением. Обычно в стандартных вычислениях рассматривается только линейная модель или одна из рассмотренных двухпараметрических зависимостей, но потом дальнейший анализ покажет отсутствие оптимумов – и не потому, что их действительно нет, а потому, что выбранная форма связи не допускает их наличия в принципе ни при каких значениях параметров.

Если же предусматривается наличие оптимумов, то следует использовать трехпараметрические формы связи, например, квадратичную модель.

Квадратичная модель должна быть принята как базовая при анализе данных с экстремумами. Обобщенная квадратичная модель с функциональными преобразованиями способная адекватно описать очень широкий класс зависимостей с экстремумами.

Построенное уравнение нелинейной регрессии дополняется показателем корреляции, а именно индексом корреляции (R):

$$R = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}},\ 0 \le R \le 1.$$

Чем ближе величина данного показателя, тем теснее связь между рассматриваемыми признаками и тем надежнее построенное уравнение регрессии.

Вопросы для самопроверки:

1) Какая цель корреляционно-регрессионного анализа?
2) Какие типы связей существуют и в чем их сущность?
3) Как располагаются сопряженные линии регрессии?
4) Какие виды регрессии различают?
5) Какие основные причины наличия в регрессионной модели случайного отклонения?
6) Как интерпретируются параметры a и b в парной линейной регрессии?
7) В чем заключается суть метода наименьших квадратов?
8) Назвать основные этапы регрессионного анализа.
9) Как связаны эмпирические коэффициенты линейной регрессии с выборочным коэффициентом корреляции между переменными уравнения регрессии?
10) Какие виды парной нелинейной регрессии различают?
11) Какие виды зависимостей можно свести функциональным преобразованием к линейным?

Тема 3. Проверка качества уравнения регрессии

3.1. Дисперсионный анализ. Коэффициент детерминации. Проверка качества построенной парной линейной модели

3.2. Оценка статистической значимости коэффициентов регрессии и корреляции

3.3. Вычисление интервалов прогноза по линейной парной регрессии

3.1. Дисперсионный анализ. Коэффициент детерминации. Проверка качества построенной парной линейной модели

Непосредственному расчету F-критерия для проверки качества построенной регрессионной модели предшествует дисперсионный анализ. Дисперсионный анализ – математический аппарат для сравнения средних нескольких популяций (групп, слоев, классов), которые определяются уровнями некоторых величин (факторов), положенных в основу классификации. Международным обозначением дисперсионного анализа является аббревиатура ANOVA (analysis of variance). В дисперсионном анализе предполагается, что группы различаются только средним уровнем результативной переменной, данные в каждой группе распределены нормально с одинаковой дисперсией.

С математической статистики известно разложение общей суммы квадратов отклонений переменной y от среднего значения $\bar{y}$ на две части – «объясненную» и «необъясненную»:

$$\sum (y-\bar{y})^2 = \sum (\hat{y}_x-\bar{y})^2 + \sum (y-\hat{y}_x)^2$$

общая сумма квадратов отклонений	=	сумма квадратов отклонений, объясненная регрессией	+ остаточная сумма квадратов отклонений

или $SSy = SSy_x + SS\varepsilon$. Эту сумму квадратов рассматривают как сумму детерминированной и случайной компонент.

Разделив каждую сумму квадратов на соответствующее число степеней свободы, получим дисперсии на одну степень свободы. Р. Фишер предложил все выкладки дисперсионного анализа оформлять в виде стандартной таблицы (m-число оцениваемых параметров уравнения регрессии при независимых переменных, n – число наблюдений) (табл. 3.1).

Таблица 3.1

Компоненты дисперсии	Число степеней свободы DF	Дисперсия на одну степень свободы	Дисперсионное отношение F	Уровень значимости
Факторная	m	$S^2{}_{y_x} = \frac{\sum (y_x - \bar{y})^2}{m}$	$F = \frac{S^2_{y_x}}{S^2_{\varepsilon}}$	α
Остаточная	$n - m - 1$	$S^2_{\varepsilon} = \frac{\sum (y - \hat{y}_x)^2}{n - m - 1}$		
Общая	$n - 1$	$S^2_y = \frac{\sum (y - \bar{y})^2}{n - 1}$		

При $F > F_{0,01}$ результативная переменная Y и классификационный показатель X не являются независимыми, они связаны между собой, между ними имеется статистическая (корреляционная) зависимость. Введем меру тесноты этой связи.

Рассматривая $SSy = SSy_x + SS\varepsilon$ как сумму детерминированной и случайной компонент, в относительных единицах имеем: $1 = \frac{SSy_x}{SSy} + \frac{SS\varepsilon}{SSy}$. Относительный вклад детерминированной части называется индексом детерминации:

$$\eta^2 = \frac{SSy_x}{SSy} = 1 - \frac{SS\varepsilon}{SSy}.$$

Индекс детерминации изменяется от 0 до 1 $(0 \le \eta^2 \le 1)$. Если индекс детерминации равен нулю, то равна нулю сумма квадратов SSy_x следовательно, равны нулю все ее члены, откуда для любой группы все средние групповые равны общему среднему, между средними групповыми нет различий, результативная переменная Y не зависит от классификационного показателя X. Если индекс детерминации равен единице, то равна нулю сумма квадратов $SS\varepsilon$, следовательно, равны нулю все ее члены, то есть никакого случайного разброса нет, каждой группе (каждому уровню классификационного показателя X) соответствует единственное значение результативной переменной $\hat{y}_x$. Чем ближе индекс детерминации к единице, тем ближе корреляционная зависимость к функциональной. Индекс детерминации η^2 показывает, какая часть полной изменчивости определяется классификационным фактором (различиями между группами наблюдений). Для совместимости с мерами тесноты связей другой природы принято извлекать корень квадратный из индекса детерминации $\eta = \sqrt{\eta^2}$. Характеристика η называется корреляционным отношением.

Для оценки качества подбора линейной функции рассчитывается коэффициент детерминации ($\eta^2 = R^2$), для парной линейной он равен $R^2 = r^2$, где $r^2 = \frac{S_{y_x}^2}{S_y^2}$.

Оценка значимости уравнения регрессии в целом выполняется с помощью F-критерия Фишера. При этом выдвигается нулевая гипотеза (H_0), что коэффициент регрессии равен нулю, т.е. $H_0: b = 0$.

Значение F-критерия Фишера для парной линейной регрессии рассчитывается $F = \frac{r^2}{1-r^2} \cdot (n-2)$. Если $F > F_{0,01}(1, n-2)$, то H_0 отвергается и делается вывод о существенности (значимости) изучаемой связи. Если $F < F_{0,05}(1, n-2)$, то H_0 не может быть отклонена без серьезного риска сделать неправильный вывод о наличии связи и уравнение регрессии считается статистически незначимым.

3.2. Оценка статистической значимости коэффициентов регрессии и корреляции

В линейной регрессии необходимо оценить значимость не только уравнения в целом, но и отдельных его параметров, поэтому определяется стандартные ошибки m_b и m_a:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_\varepsilon^2}{\sum (x - \bar{x})^2}};$$

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n-2} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S_\varepsilon^2 \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}}.$$ Вычисляются t-значения Стьюдента: $t_b = \frac{b}{m_b}$; $t_a = \frac{a}{m_a}$ и сравниваются с табличным при числе

степеней свободы $n-2$. Если $t_b > t_{\alpha=0{,}05}(n-2)$, то гипотезу о несущественности коэффициента регрессии можно отклонить.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции m_r:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}, \quad t_r = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}.$$

Следовательно $t_r^2 = F$, а также $t_b^2 = F$, а значит $t_r^2 = t_b^2$. Т.е. проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

Зная значения стандартных ошибок m_b и m_a, можно вычислить доверительные интервалы для параметров уравнения регрессии:

$b \pm t_{\alpha,\, n-2} m_b$; $a \pm t_{\alpha,\, n-2} m_a$.

3.3. Вычисление интервалов прогноза по линейной парной регрессии

Одной из основных задач построения регрессионной модели является использование ее для вычисления прогноза. Прогноз результативного показателя будет реальным, если вычислить доверительный интервал для значения переменной y при заданном значении переменной x:

$$\hat{y}_x - m_{\hat{y}_x} \le y_p \le \hat{y}_x + m_{\hat{y}_x},$$

где $m_{\hat{y}_x}$ - стандартная ошибка.

Доказывается, что стандартная ошибка $m_{\hat{y}_x}$ предсказываемого среднего значения y при заданном значении x_0 рассчитывается по формуле [31, с. 57 – 60]:

$$m_{\hat{y}_x} = S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2}},$$

где S - стандартная ошибка, которая вычисляется как $S = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-2}}$.

Величина стандартной ошибки $m_{\hat{y}_x}$ достигает минимума при $\bar{x}$ и возрастает по мере удаления от него. Вычисление прогнозируемого значения $\hat{y}_x$ с 95%-ным доверительным интервалом выполняется по формуле $\hat{y}_{x_p = x_0} \pm t_\alpha \cdot m_{\hat{y}_x}$.

Средняя ошибка прогнозируемого индивидуального значения $y(x_k)$ составит:

$$m_{y(x_k)} = S\sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2}}$$

На графике доверительные границы для $\hat{y}_x$ представляют собой гиперболы, расположенные по обе стороны от линии регрессии.

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации как среднюю арифметическую простую:

$$A = \frac{1}{n}\sum\left|\frac{(y - \hat{y}_x)}{y}\right| \cdot 100 .$$

В анализе моделей влияния фактора на результат полезным является расчет коэффициента эластичности: $Э = f'(x)\dfrac{x}{y}$, где $f'(x)$ - первая производная, характеризующая соотношение приростов результата и фактора для соответствующей формы связи. Поскольку коэффициент эластичности для линейной функции не является величиной постоянной, а зависит от соответствующего значения *x*, то рассчитывается средний показатель эластичности по формуле: $\bar{Э} = b\dfrac{\bar{x}}{\bar{y}}$.

Пример. 1) Проверить гипотезу о наличии линейной связи между переменными: себестоимостью единицы изделия (*y*, тыс. грн) и величиной выпуска продукции (*x*, тыс. шт.) (продолжение примера лекции 2) в генеральной совокупности. Доверительная вероятность $p = 95\%, \quad n = 5$.

Решение. Имеем H_0: $r_г = 0$, т.е. между переменными отсутствует линейная связь в генеральной совокупности; H_1: $r_г \neq 0$, т.е. между переменными есть линейная связь в генеральной совокупности. $\alpha = 0{,}05$, было рассчитано, что $r = -0{,}904$, поэтому

$$t = \sqrt{\frac{r^2(n-2)}{1-r^2}} = \sqrt{\frac{0{,}817(5-2)}{1-0{,}817}} = 3{,}66; \qquad t > t_{\alpha=0{,}05,\,3}; \quad 3{,}66 > 3{,}1825 .$$

Отклоняем гипотезу H_0 и принимаем гипотезу H_1 на уровне значимости 5%.

Между себестоимостью единицы изделия (y, тыс. грн) и величиной выпуска продукции (x, тыс. шт.) есть линейная связь в генеральной совокупности.

2) Найти доверительный интервал для коэффициента регрессии зависимости себестоимости единицы изделия (y, тыс. грн) и величины выпуска продукции (x, тыс. шт.) при $p = 95\%, \quad n = 5$.

Решение.

$b \pm t_{\alpha,\, n-2} m_b$;

Номер	x	y	$(x-\bar{x})^2$	$(y-\hat{y})^2$
1	2	1,9	4	0,00
2	3	1,7	1	0,0081
3	4	1,8	0	0,0144
4	5	1,6	1	0,0009
5	6	1,4	4	0,0036
Сумма	20	8,4	10	0,027

$$m_b = \sqrt{\frac{\sum (y-\hat{y}_x)^2 / (n-2)}{\sum (x-\bar{x})^2}} = \sqrt{\frac{S_\varepsilon^2}{\sum (x-\bar{x})^2}} = \sqrt{\frac{0{,}027/3}{10}} = 0{,}03$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{20}{5} = 4; \; t_{\alpha=0{,}05,\,3} = 3{,}182$$

$b \pm t_{\alpha,\, n-2} m_b = -0{,}11 \pm 3{,}182 \cdot 0{,}03 = -0{,}11 \pm 0{,}1$, т.е. $-0{,}21 < b < -0{,}01$.

3) Найти доверительный интервал для среднего значения переменной y при заданном значении $x_0 = 5{,}5$ тыс. шт. Доверительная вероятность $p = 95\%, \quad n = 5$.

Решение

Номер	x	x^2	$(y-\hat{y})^2$
1	2	4	0,00
2	3	9	0,0081
3	4	16	0,0144
4	5	25	0,0009
5	6	36	0,0036
Сумма	20	90	0,027

$$y = a + bx_0 \pm t_{\alpha, n-2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}, \; S = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-2}};$$

$$\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 90 - \frac{400}{5} = 10; \; S = \sqrt{\frac{0{,}027}{3}} = 0{,}0949;$$

$$(x_0 - \bar{x})^2 = (5{,}5 - 4)^2 = 2{,}25;$$

$$y = 2{,}12 + (-0{,}11)5{,}5 \pm 3{,}182 \cdot 0{,}0949 \sqrt{\frac{1}{5} + \frac{2{,}25}{10}} = 1{,}515 \pm 0{,}197.$$

Т.е. доверительный интервал для среднего значения переменной y при задан ном значении $x_0 = 5{,}5$ тыс.шт. равен $1{,}318 < y < 1{,}712$, т.е. $(1{,}318; 1{,}712)$.

4) Найти доверительный интервал для индивидуального значения переменной y при заданном значении $x_0 = 5{,}5$ тыс. шт. Доверительная вероятность $p = 95\%$, $n = 5$.

Решение

$$a + bx_0 \pm t_{\alpha, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}},$$

$$2{,}12 + (-0{,}11)5{,}5 \pm 3{,}182 \cdot 0{,}0949 \sqrt{1 + \frac{1}{5} + \frac{2{,}25}{10}} = 1{,}515 \pm 0{,}3605.$$

Т.е. доверительный интервал для индивидуального значения переменной y при заданном значении $x_0 = 5{,}5$ тыс.шт. равен $1{,}1545 < y < 1{,}8755$, т.е. $(1{,}1545; 1{,}8755)$. Доверительный интервал для индивидуальных значений переменной y шире доверительного интервала для среднего значения переменной y (при заданной значении x_0)

Вопросы для самопроверки:

1) Какая основная идея дисперсионного анализ?

2) Как проверить качество построенной парной линейной модели с помощью критерия Фишера?
3) Как рассчитывается и что показывает индекс детерминации?
4) Как оценить значимость коэффициентов регрессии?
5) Как оценить значимость коэффициента корреляции?
6) Как построить доверительные интервалы для параметров уравнения регрессии
7) Как построить доверительный интервал для значения переменной y при заданном значении переменной x?
8) Можно ли определить качество модели из относительных отклонений по каждому наблюдению?
9) Какое предназначение коэффициента эластичности в регрессионном анализе?

Тема 4. Линейные модели множественной регрессии

4.1. Общие вопросы построения множественной регрессионной модели
4.2. Матричная форма регрессионного анализа
4.3. Регрессионная модель в стандартизованных переменных
4.4. Коэффициенты частной корреляции

4.1. Общие вопросы построения множественной регрессионной модели

В практической деятельности учитывают влияние не одного фактора на результат, а нескольких. Так, на рентабельность производства влияют и производительность труда, и материалоемкость, и фондоотдача, и оборачиваемость оборотных средств, и ритмичность производства. Поэтому чаще всего для анализа или для прогнозирования необходимо учитывать много факторов, то есть необходимо разработать множественную регрессионную модель.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на результативный показатель.

Построение модели начинается с определения ее спецификации. Суть проблемы спецификации модели предполагает решение отбора факторов и выбора вида уравнения регрессии. Факторы должны быть количественными, не коррелировать между собой и не быть функционально зависимыми. Отбор выполняется на основе теоретико-экономического анализа.

Исходя из четкой интерпретации параметров, наиболее широко используются в экономике линейная и степенная зависимости. Уравнение линейной множественной регрессии $y = a + b_1 x_1 + b_2 x_2 + ... + b_m x_m + \varepsilon$ (уравнение «чистой» регрессии). Коэффициенты «чистой» регрессии b_i характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменной значении других факторов, закрепленных на среднем уровне. В степенной функции $\hat{y}_x = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot ... \cdot x_m^{b_m}$ коэффициенты b_i (коэффициенты эластичности) показывают на сколько процентов изменяется в среднем результат с изменением соответствующего фактора на 1% при неизменности действия других факторов. Этот вид уравнения регрессии получил наибольшее распространение в производственных функциях, в исследованиях спроса и потребления.

Параметры уравнения множественной регрессии оцениваются, как и в парной регрессии, методом наименьших квадратов (МНК). При его применении строится система нормальных уравнений, решая которую получают оценки параметров уравнения. Для уравнения $y = a + b_1 x_1 + b_2 x_2 + ... + b_m x_m + \varepsilon$ система нормальных уравнений будет:

$$\begin{cases} \sum y = n \cdot a + b_1 \cdot \sum x_1 + b_2 \cdot \sum x_2 + ... + b_m \cdot \sum x_m \\ \sum y \cdot x_1 = a \cdot \sum x_1 + b_1 \cdot \sum x_1^2 + b_2 \cdot \sum x_1 x_2 + ... + b_m \cdot \sum x_1 x_m \\ ... \quad ... \quad ... \quad ... \quad ... \quad ... \quad ... \quad ... \\ \sum y \cdot x_m = a \cdot \sum x_m + b_1 \cdot \sum x_1 \cdot x_m + b_2 \cdot \sum x_2 x_m + ... + b_m \cdot \sum x_m^2 \end{cases}$$

Ее решение можно найти по правилу Крамера: $a = \frac{\Delta a}{\Delta}, b_1 = \frac{\Delta b_1}{\Delta}, ..., b_m = \frac{\Delta b_m}{\Delta}$, где $\Delta -$ определитель системы, при этом

$$\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & ... & \sum x_m \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & ... & \sum x_m x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & ... & \sum x_m x_2 \\ ... & ... & ... & ... & ... \\ \sum x_m & \sum x_1 x_m & \sum x_2 x_m & ... & \sum x_m^2 \end{vmatrix} \neq 0,$$

$\Delta a, \Delta b_1, ..., \Delta b_m -$ частные определители, получаемые путем замены соответствующего столбца определителя системы данными левой части.

4.2. Матричная форма регрессионного анализа

Обозначим столбцы значений переменных (экономических показателей) y, x_i, ε буквами Y, X_i, E. Запишем уравнение регрессии

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + \varepsilon$$

для всех наблюдений переменных: $a = b_0$

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{pmatrix} b_0 + \begin{pmatrix} x_{11} \\ x_{12} \\ \ldots \\ x_{1n} \end{pmatrix} b_1 + \begin{pmatrix} x_{21} \\ x_{22} \\ \ldots \\ x_{2n} \end{pmatrix} b_2 + \ldots + \begin{pmatrix} x_{m1} \\ x_{m2} \\ \ldots \\ x_{mn} \end{pmatrix} b_m + \begin{pmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{pmatrix},$$

где обозначим $X = (X_0\ X_1\ X_2 \ldots X_m) = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{m1} \\ 1 & x_{12} & x_{22} & \ldots & x_{m2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{mn} \end{pmatrix}$,

$Y' = (y_1, y_2, \ldots, y_n)$,

$X_0 = (1, 1, \ldots, 1)$, $B = (b_0, b_1, b_2, \ldots, b_m)$, $E = (e_1, e_2, \ldots, e_n)$.

Тогда в векторной форме равенства пишутся как:

$$Y = X_0 b_0 + X_1 b_1 + X_2 b_2 + \ldots + X_m b_m + E,$$

а в матричной форме: $Y = XB + E$. Умножим это равенство на матрицу X' и учтем, что $X_i'E = 0$, получим в матричной форме систему нормальных уравнений:

$$X'Y = X'X\,B.$$

Откуда находим вектор коэффициентов регрессии в виде:

$$B = (X'X)^{-1}X'Y = CX'Y,\ B = CX'Y$$

где $C = (X'X)^{-1}$ – матрица, обратная к матрице системы нормальных уравнений $X'X$, т.е. $C(X'X) = I$ (единичная матрица).

4.3. Регрессионная модель в стандартизованных переменных

Для определения рейтинга влияния факторов в модели вычисляют регрессионную модель в стандартизированных переменных. Все формулы регрессионного анализа в стандартизованных переменных приобретают вид намного проще. Обозначим $t_y = \dfrac{y - \bar{y}}{s_y}$, $t_{x_i} = \dfrac{x_i - \bar{x}_i}{s_{x_i}}$.

Тогда $y - \bar{y} = b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + \ldots + b_m(x_m - \bar{x}_m)$,

$$\frac{y-\bar{y}}{s_y}=\beta_1\frac{(x_1-\bar{x}_1)}{s_{x_1}}+\beta_2\frac{(x_2-\bar{x}_2)}{s_{x_2}}+\ldots+\beta_m\frac{(x_m-\bar{x}_m)}{s_{x_m}}+\frac{e}{s_y},$$

$$t_y=\beta_1 t_1+\beta_2 t_2+\ldots+\beta_m t_m+\varepsilon,\ \varepsilon=\frac{e}{s_y},$$

Коэффициент «чистой» регрессии связаны со стандартизованными коэффициентами регрессии: $\beta_i=b_i\frac{s_{x_i}}{s_y}$, $a=\bar{y}-b_1\bar{x}_1-b_2\bar{x}_2-\ldots-b_m\bar{x}_m$

Применяя МНК к уравнению множественной регрессии в стандартизованном масштабе, после преобразований получаем систему нормальных уравнений:

$$\begin{cases} r_{yx_1}=\beta_1+\beta_2 r_{x_1x_2}+\ldots+\beta_m r_{x_1x_m}\\ r_{yx_2}=\beta_1 r_{x_2x_1}+\beta_2+\ldots+\beta_m r_{x_2x_m}\\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\\ r_{yx_m}=\beta_1 r_{x_mx_1}+\beta_2 r_{x_mx_2}+\ldots+\beta_m \end{cases},$$

Стандартизованные коэффициенты регрессии показывают, на сколько сигм (среднеквадратических отклонений) изменится в среднем результат, если соответствующий фактор x_i изменится на одну сигму при неизменном среднем уровне других факторов. Поскольку переменные центрированные и нормированные, стандартизованные коэффициенты регрессии β_i сравнимы между собой и поэтому можно ранжировать факторы по силе их воздействия на результат.

4.4. Множественная корреляция

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, или оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции:

$$R_{yx_1\ldots x_m}=\sqrt{1-\frac{S_\varepsilon^2}{S_y^2}},$$

где S_y^2- общая дисперсия результативного признака; S_ε^2- остаточная дисперсия. $0\le R_{yx_1\ldots x_m}\le 1$, чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов.

При линейной зависимости признаков формула индекса корреляции может быть представлена в виде: $R=\sqrt{\beta_{x_1}\cdot r_{yx_1}+\beta_{x_2}\cdot r_{yx_2}+\ldots+\beta_{x_m}\cdot r_{yx_m}}=\sqrt{\sum\beta_{x_i}\cdot r_{yx_i}}$.

При линейной зависимости линейный коэффициент множественный корреляции определяется:

$$R_{yx_1\dots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

где $\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_mx_1} & r_{x_2x_m} & \dots & 1 \end{vmatrix}$; $\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_2x_m} & \dots & 1 \end{vmatrix}$.

Ранжирование факторов по силе влияния на результат можно провести через β-коэффициенты или же с помощью частных коэффициентов корреляции (для линейных связей) и частные индексы детерминации (при нелинейных связях).

Частные коэффициенты (индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, $r_{yx_1 \cdot x_2}$- коэффициент частной корреляции первого порядка. Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле:

$$r_{yx_i \cdot x_1x_2\dots x_m} = \frac{r_{yx_i \cdot x_1x_2\dots x_{m-1}} - r_{yx_m \cdot x_1x_2\dots x_{m-1}} \cdot r_{x_ix_m \cdot x_1x_2\dots x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_m \cdot x_1x_2\dots x_{m-1}}\right)\left(1 - r^2_{x_ix_m \cdot x_1x_2\dots x_{m-1}}\right)}}.$$

При двух факторах и $i = 1$ данная формула примет вид: $r_{yx_1 \cdot x_2} = \dfrac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right)\left(1 - r^2_{x_1x_2}\right)}}$. Соответственно при $i = 2$ и двух факторах частный коэффициент корреляции y с фактором x_2 можно определить по формуле $r_{yx_2 \cdot x_1} = \dfrac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right)\left(1 - r^2_{x_1x_2}\right)}}$. Существует взаимосвязь между β-коэффициентами и коэффициентами корреляции:

$$\begin{cases} \beta_{x_1} = \dfrac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} \\ \beta_{x_2} = \dfrac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} \end{cases}$$

В эконометрике частные коэффициенты корреляции обычно не имеют самостоятельного значения. В основном их используют на стадии формирования модели, в частности в процедуре отсева факторов.

Пример. В процессе изучения зависимости прибыли (тыс. грн) y от выработки продукции на одного работника (ед.) x_1 и индекса цен на продукцию (%) x_2 получены данные по 30 предприятиям

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	250	38	$r_{yx_1} = 0{,}68$
x_1	47	12	$r_{yx_2} = 0{,}63$
x_2	112	21	$r_{x_1x_2} = 0{,}42$

1) Построить уравнение множественной регрессии в стандратизованной форме и уравнение «чистой» регрессии.

2) Рассчитать линейные коэффициенты частной корреляции и коэффициент множественной корреляции, сравнить их с линейными коэффициентами парной корреляции, пояснить различия между ними.

Решение

1) $$\beta_{x_1} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0{,}68 - 0{,}63 \cdot 0{,}42}{1 - 0{,}42^2} = 0{,}5044;$$

$$\beta_{x_2} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0{,}63 - 0{,}68 \cdot 0{,}42}{1 - 0{,}42^2} = 0{,}4182.$$

Получим уравнение $t_y = 0{,}5044 t_{x_1} + 0{,}4182 t_{x_2}$.

Таким образом, первый фактор – выработка продукции на одного работника (ед.) x_1 влияет больше на изменение прибыли (тыс. грн) y, по сравнению с индексом цен на продукцию (%) x_2.

Для построения уравнения «чистой» регрессии рассчитаем b_1 и b_2, используя формулы для перехода от β_i к b_i:

$$\beta_i = b_i \frac{s_{x_i}}{s_y}; \quad b_i = \beta_i \frac{s_y}{s_{x_i}}.$$

$$b_1 = 0{,}5044\frac{38}{12} = 1{,}5973; \quad b_2 = 0{,}4182\frac{38}{21} = 0{,}7567.$$

Значение a определяем из соотношения

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 = 250 - 1{,}5973 \cdot 47 - 0{,}7567 \cdot 21 = 159{,}0362;$$

$$\hat{y}_x = 159{,}0362 + 1{,}5973x_1 + 0{,}7567x_2.$$

Таким образом, с изменением выработки продукции на одного работника (ед.) x_1 на единицу при неизменной значении второго фактора, закрепленного на среднем уровне, среднее изменение прибыли составит 1,5973 тыс. грн, а с изменением индекса цен на продукцию на 1% (x_2) среднее изменение прибыли составит 0,7567 тыс. грн.

2) Коэффициенты частной корреляции:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}} = \frac{0{,}68 - 0{,}63 \cdot 0{,}42}{\sqrt{(1 - 0{,}63^2)(1 - 0{,}42^2)}} = 0{,}5894;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})}} = \frac{0{,}63 - 0{,}68 \cdot 0{,}42}{\sqrt{(1 - 0{,}68^2)(1 - 0{,}42^2)}} = 0{,}5176.$$

Ввиду имеющейся межфакторной связи $r_{x_1x_2} = 0{,}42$, наблюдаем расхождение между значениями коэффициентами парной и частной корреляции:

$r_{yx_1} = 0{,}68$; $\qquad r_{yx_2} = 0{,}63$;

$r_{yx_1 \cdot x_2} = 0{,}5894$; $\qquad r_{yx_2 \cdot x_1} = 0{,}5176$.

Выводы о тесноте и направлении связи на основе коэффициентов парной и частной корреляции совпадают.

Коэффициент множественной корреляции равен:

$$R = \sqrt{\beta_{x_1} \cdot r_{yx_1} + \beta_{x_2} \cdot r_{yx_2}} = \sqrt{0{,}5044 \cdot 0{,}68 + 0{,}4182 \cdot 0{,}63} = \sqrt{0{,}606456} = 0{,}77875.$$

Зависимость y от x_1 и x_2 характеризуется как тесная, в которой 60% вариации средней прибыли (тыс. грн) определяются вариацией выработкой продукции на одного работника (ед.) и индексом цен на продукцию (%). Прочие факторы, не включенные в модель, составляют соответственно 40% от общей вариации y.

Вопросы для самопроверки:

1) В чем суть спецификации множественной регрессионной модели?
2) Как оцениваются параметры множественной линейной регрессионной модели?

3) Какая цель построения регрессионной модели в стандартизованных переменных?
4) Какая связи между коэффициентами множественной линейной «чистой» регрессии и в стандартизованных переменных?
5) Как интерпретируются β-коэффициенты?
6) Как вычисляется индекс множественной линейной корреляции?
7) Какое основное предназначение коэффициентов частной корреляции?

Тема 5. Оценка надежности общей многофакторной линейной модели

5.1. Проверка общего качества уравнения регрессии.

5.2. Проверка статистической значимости коэффициентов уравнения регрессии

5.1. Проверка общего качества уравнения регрессии

Значимость уравнения множественной регрессии в целом оценивается с помощью F-критерия Фишера. Проверяется гипотеза H_0 о статистической значимости коэффициента детерминации ($H_0: R^2 = 0$). Для проверки данной гипотезы используется F-статистика:

$$F = \frac{D_{\text{факт}}}{D_{\text{ост}}} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m},$$

где $D_{\text{факт}}$ - факторная сумма квадратов на одну степень свободы;

$D_{\text{ост}}$ - остаточная сумма квадратов на одну степень свободы;

R^2- коэффициент детерминации; m- число параметров при переменных (в линейной регрессии совпадает с числом включенных в модель факторов);

n - число наблюдений.

Величина F при выполнении предпосылок МНК и при справедливости H_0 имеет распределение Фишера, аналогичное распределению F-статистики.

Показатели F и R^2 равны или не равны нулю одновременно. Если F =0, то R^2=0, и линия регрессии $y = \bar{y}$ является наилучшей по МНК, и, следовательно, переменная y линейно не зависит от $x_1, x_2, ..., x_m$. Для проверки нулевой гипотезы при заданном уровне значимости α по таблицам критических точек распределения Фишера находится критическое значение

$F_{кр} = F_{\alpha}(m, n-m-1)$. Нулевая гипотеза отклоняется, если $F > F_{кр}$. Это равносильно тому, что $R^2 > 0$, т.е. R^2 статистически значим.

5.2. Проверка статистической значимости коэффициентов уравнения регрессии

Ввиду корреляции между факторами значимость одного и того же фактора может быть различной в зависимости от последовательности его введения в модель. Мерой для оценки включения фактора в модель служит частный F-критерий Фишера. Он построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом:

$$F_{x_1} = \frac{R^2_{yx_1x_2...x_m} - R^2_{yx_2...x_m}}{1 - R^2_{yx_1x_2...x_m}} \cdot \frac{n-m-1}{1},$$

где $R^2_{yx_1x_2...x_m}$ - коэффициент детерминации для модели с полным набором факторов; $R^2_{yx_2...x_m}$ - тот же показатель, но без включения в модель фактора x_1.

В числителе $R^2_{yx_1x_2...x_m} - R^2_{yx_2...x_m}$ - прирост доли объясненной вариации y за счет дополнительного включения в модель фактора x_1.

В общем виде для фактора x_i частный F-критерий определяется как:

$$F_{x_i} = \frac{R^2_{yx_i x_1 x_2..x_i...x_m} - R^2_{yx_1,...,x_{i-1},x_{i+1},...,x_m}}{1 - R^2_{yx_1x_2..x_i...x_m}} \cdot \frac{n-m-1}{1}.$$

В числителе $R^2_{yx_i x_1 x_2..x_i...x_m} - R^2_{yx_1,...,x_{i-1},x_{i+1},...,x_m}$ - прирост доли объясненной вариации y за счет дополнительного включения в модель фактора x_i.

Когда в модели два фактора имеем:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-m-1}{1}; \quad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-m-1}{1}.$$

Частный F-критерий Фишера оценивает значимость коэффициентов чистой регрессии. Зная величину F_{x_i} можно определить и t-критерий для коэффициента регрессии при i-м факторе: $t_{b_i} = \sqrt{F_{x_i}}$. Если расчетное значение t_{b_i} больше табличного, то подтверждается значимость включенного в модель фактора.

Оценка значимости коэффициентов чистой регрессии по t-критерию Стьюдента может быть проведена без расчетов частных F-критериев:

$$t_{b_i} = \frac{b_i}{m_{b_i}},$$

где
$$m_{b_i} = \frac{S_y\sqrt{1-R^2_{yx_1\ldots x_m}}}{S_{x_i}\sqrt{1-R^2_{x_i x_1\ldots x_m}}}\cdot\frac{1}{\sqrt{n-m-1}};$$

$$b_i = \frac{S_y\sqrt{R^2_{yx_1\ldots x_m} - R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}}{S_{x_i}\sqrt{1-R^2_{x_i x_1\ldots x_m}}}.$$

Взаимосвязь частного коэффициента корреляции, частного F-критерия и t-критерия Стьюдента для коэффициентов чистой регрессии используется в процедурах отбора факторов. Отбор факторов методом исключения осуществляется не только по частным коэффициентам корреляции, исключая фактор с наименьшим значением частного коэффициента корреляции, но и по величинам t_{b_i} и F_{x_i}. Частный F-критерий широко используется и при построении модели методами включения и пошагового отбора переменных.

Пример. **(**Условие примера в лекции 4) Рассчитать общий и частные F-критерий Фишера.

$$F_p = \frac{R^2}{1-R^2}\cdot\frac{n-m-1}{1} = \frac{0{,}6065}{1-0{,}6065}\cdot\frac{30-2-1}{1} = 41{,}615;\ F_{\alpha=0{,}05}(1;27)=4{,}21.$$

$F_p > F_{табл}$. Следовательно, уравнение зависимости прибыли (тыс. грн) y от выработки продукции на одного работника (ед.) x_1 и индекса цен на продукцию (%) x_2 статистически значимо.

Проверим значимость коэффициентов регрессии в уравнении множественной регрессии. Частный F-критерий для фактора x_1 определим по формуле:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1-R^2_{yx_1x_2}}\cdot\frac{n-m-1}{1} = \frac{0{,}6065-0{,}63^2}{1-0{,}6065}\cdot\frac{30-2-1}{1} = 14{,}3817.$$

Табличное значение равно $F_{\alpha=0{,}05}(1;27)=4{,}21$. Коэффициент регрессии в модели статистически значим. Включение в модель фактора x_1 после фактора x_2 статистически оправдано – доля объясненной вариации возросла на $(0{,}6065-0{,}63^2)\cdot 100 = 20{,}96\%$.

Частный F-критерий для фактора x_2 определим по формуле:

$$F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-m-1}{1} = \frac{0{,}6065 - 0{,}68^2}{1 - 0{,}6065} \cdot \frac{30-2-1}{1} = 9{,}8824,$$ что больше табличного. Коэффициент регрессии в модели статистически значим. Включение в модель фактора x_2 после фактора x_1 статистически оправдано – доля объясненной вариации возросла на $(0{,}6065 - 0{,}68^2) \cdot 100 = 14{,}41\%$.

Пример. Определить статистическую значимость коэффициентов линейной множественной регрессионной модели зависимости объема предложения товара от цены товара x_1 и зарплаты сотрудников x_2:

$$\hat{y} = 90{,}74 + 0{,}88x_1 - 7{,}32x_2,$$

если $n = 10$; $m_a = 24{,}87$; $m_{b_1} = 0{,}36$; $m_{b_2} = 1{,}91$. Доверительная вероятность 95%.

Обозначим $H_0 : b_i = 0$, то есть объясняющая переменная x_i не влияет на результативный признак y.

$H_1 : b_i \neq 0$, то есть объясняющая переменная x_i влияет на результативный признак y.

Вычислим значения t-статистики:

$$t_a = \frac{a}{m_a} = \frac{90{,}74}{24{,}87} = 3{,}649;$$

$$t_{b_1} = \frac{b_1}{m_{b_1}} = \frac{0{,}88}{0{,}36} = 2{,}44;$$

$$t_{b_2} = \frac{b_2}{m_{b_2}} = -\frac{7{,}32}{1{,}91} = -3{,}832.$$

Поскольку граничные точки $\pm t_{\alpha,\, n-m-1} = \pm 2{,}365$, отвергаем гипотезу H_0 и принимаем H_1 на уровне значимости 5%. Все коэффициенты статистически значимы, то есть x_1 и x_2 влияют на результативный признак y.

Вопросы для самопроверки:

1) Как использовать критерий Фишера для проверки общего качества уравнения регрессии?
2) Какие существуют методы проверки статистической значимости коэффициентов уравнения регрессии?
3) Существует ли связь между методами проверки статистической значимости коэффициентов уравнения регрессии?

Тема 6. Мультиколлинеарность, ее последствия и методы устранения

6.1. Предпосылки метода наименьших квадратов.

6.2. Суть мультиколлинеарности. Последствия мультиколлинеарности.

6.3. Определение мультиколлинеарности. Методы устранения мультиколлинеарности.

6.1. Предпосылки метода наименьших квадратов

Самым распространенным методом оценки параметров уравнения множественной линейной регрессии является метод наименьших квадратов. Свойства оценок коэффициентов регрессии, качество построенной регрессии существенно зависят от свойств случайной составляющей. Доказано, что для получения по МНК наилучших результатов необходимо выполнение предпосылок относительно случайного отклонения.

***Предпосылки метода наименьших квадратов* (*условия Гаусса-Маркова*):**

1. Математическое ожидание случайного отклонения ε_i равно нулю для всех наблюдений: $M(\varepsilon_i) = 0, i = \overline{1,n}$.

Данное условие означает, что случайное отклонение в среднем не оказывает влияния на зависимую переменную.

2. Гомоскедастичность (постоянство дисперсии отклонений). Дисперсия случайных отклонений ε_i постоянна: $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$ для любых наблюдений i и j.

Данное условие подразумевает: несмотря на то, что при каждом конкретном наблюдении случайное отклонение может быть либо большим, либо меньшим, не должно быть никакой априорной причины, вызывающей большую ошибку (отклонение).

3. Отсутствие автокорреляции.

Случайные отклонения ε_i и ε_j являются независимыми друг от друга для всех $i \neq j$:

$$\sigma_{\varepsilon_i \varepsilon_j} = \operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0 & \text{если } i \neq j \\ \sigma^2 & \text{если } i = j \end{cases}.$$

С учетом выполнимости предпосылки 1 данное соотношение можно записать в виде $M(\varepsilon_i \varepsilon_j) = 0 \ (i \neq j)$.

4. Случайное отклонение должно быть независимо от объясняющих переменных: $\sigma_{\varepsilon_i x_j} = 0$.

Обычно это условие выполняется автоматически, если объясняющие переменные не являются случайными в данной модели.

5. Модель является линейной относительно параметров.

Для случая множественной линейной регрессии существенными являются еще две предпосылки.

6. Отсутствие мультиколлинеарности.

Между объясняющими переменными отсутствует строгая (сильная) линейная зависимость.

7. Ошибки $\varepsilon_i, i = \overline{1,n}$ имеют нормальное распределение ($\varepsilon_i \sim N(0,\sigma)$).

Выполнимость данной предпосылки важна для проверки статистических гипотез и построение интервальных оценок.

Теорема Гаусса-Маркова. Если предпосылки 1 – 5 выполнены, то оценки полученные по МНК, обладают следующими свойствами:

1. Оценки являются несмещенными. Оценка b генеральной характеристики β называется несмещенной, если $M(b_i) = \beta_i$. Несмещенные оценки не имеют систематических смещений.
2. Оценки состоятельны, поскольку дисперсия оценок параметров при возрастании числа n наблюдений стремится к нулю: $D(a) \xrightarrow[n\to\infty]{} 0$, $D(b_i) \xrightarrow[n\to\infty]{} 0$.
3. Оценки эффективны, то есть они имеют наименьшую дисперсию по сравнению с любыми другими оценками данных параметров, линейны относительно величин y_i.

Таким образом, при выполнении предпосылок МНК относительно ошибок ε_i оценки $b_0, b_1, \dots, b_m$ параметров множественной линейной регрессии по МНК являются несмещенными, эффективны и состоятельными, т.е. *BLUE*-оценками (*Best Linear Unbiased Estimators*) – наилучшими линейными несмещенными оценками.

6.2. Суть мультиколлинеарности. Последствия мультиколлинеарности.

Серьезной проблемой при построении моделей множественной линейной регрессии по МНК является мультиколлинеарность – линейная взаимосвязь двух или нескольких объясняющих переменных. Мультиколлинеарность может быть проблемой лишь в случае множественной регерессии.

Причиной возникновения мультиколлинеарности в экономических исследованиях является существование соотношений между объясняющими переменными. Это касается регрессии, построенной как на результатах одновременных обследований, так и по данным, полученным из временных рядов. Например, при исследовании зависимости затрат от объема производства и введенных в действие основных фондов необходимо учесть, что объем производства зависит также от основных фондов. Поэтому коэффициенты регрессии не будут точно отражать зависимость от данных двух факторов, так как основные фонды оказывают также влияние на затраты через объем производства.

Система нормальных уравнений хотя и имеет решение, поскольку определитель матрицы $X'X$ отличен от нуля и матрица $X'X$ невырожденная, но имеют место большие стандартные ошибки. Чем сильнее корреляция между факторами, тем меньше определитель матрицы $X'X$. Это приводит к значительному понижению точности оценки параметров регрессии, искажению оценок дисперсии остатков, дисперсии коэффициентов регрессии и ковариации между ними.

Обычно выделяются следующие последствия мультиколлинеарности [2, с. 271 – 281]:

1) большие дисперсии (стандартные ошибки) оценок. Это затрудняет нахождение истинных значений определяемых величин и расширяет интервальные оценки, ухудшая их точность.

2) уменьшаются t-статистики коэффициентов, что может привести к неоправданному выводу о существенности влияния соответствующей объясняющей переменной на зависимую переменную.

3) оценки коэффициентов по МНК и их стандартные ошибки становятся очень чувствительными к малейшим изменениям данных, т.е. они становятся неустойчивыми;
4) затрудняется определение вклада каждой из объясняющих переменных в объясняемую уравнением регрессии дисперсию зависимой переменной.
5) возможно получение неверного знака у коэффициента регрессии.

6.3. Определение мультиколлинеарности. Методы устранения мультиколлинеарности.

Существует несколько признаков, по которым может быть установлено наличие мультиколлинеарности:
1) коэффициент детерминации R^2 достаточно высок, но некоторые из коэффициентов регрессии статистически незначимы, т.е. они имеют низкие t-статистики;
2) парная корреляция между малозначимыми объясняющими переменными достаточно высока;
3) высокие частные коэффициенты корреляции;

Частные коэффициенты корреляции определяют силу линейной зависимости между двумя переменными без учета влияния на них других переменных. Однако при изучении многомерных связей в ряде случаев парные коэффициенты корреляции могут давать совершенно неверные представления о характере связи между двумя переменными. Это объясняется тем, что обе переменные изменяются в одном направлении под влиянием других переменных, как учтенных в модели, так и неучтенных. Поэтому необходимо измерять действительную силу линейной связи между двумя переменными, очищенную от влияния на рассматриваемую пару переменных других факторов.
4) сильная вспомогательная (дополнительная регрессия), т.е. какая-либо из объясняющих переменных является линейной комбинацией других объясняющих переменных.

Мультиколлинеарность может иметь место вследствие того, что одна из объясняющих переменных является линейной комбинацией других объясняющих переменных. Для анализа строятся уравнения регрессии каждой из объясняющих переменных от оставшихся. Вычисляются соответствующие коэффициенты детерминации R_j^2 и рассчитывается их статистическая значимость по критерию Фишера. Если коэффициент R_j^2 статистически незначим, то соответствующая переменная не является линейной комбинацией других переменных и ее можно оставить в уравнении регрессии. В противном случае есть основания считать, что эта переменная существенно зависит от других объясняющих переменных и имеет место мультиколлинеарность.

Единого метода устранения мультиколлинеарности не существует, но имеются часто используемые рекомендации [31, с. 214 – 222]:

1) исключение переменной из модели. Этот метод заключается в том, что высоко коррелирующие объясняющие переменные исключаются из регрессии, и она заново оценивается. Практикой доказано, что если $|r_{ij}| > 0{,}8$, то одну из переменных следует исключить. Какую именно переменную необходимо исключить определяют на основании экономического анализа зависимой переменной.
2) исключение тренда. При построении регрессии по данным, полученным из временных рядов, рекомендуется исключить тренд или компенсировать изменение последовательных значений переменных (прирост). Этим достигается соблюдение предпосылок регрессионного анализа – независимость наблюдений и уменьшение мультиколлинеарности.
3) получение дополнительных данных или новой выборки;
4) изменение спецификации модели: либо изменяется форма модели, либо добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на зависимую переменную;
5) использование предварительной информации о некоторых параметрах. Обычно на основе ранее построенных регрессионных уравнений или проведенных экономических исследований уже представление о величине или соотношении двух или нескольких коэффициентов регрессии. Такой предварительной информацией пользуются при построении регрессионной модели.
6) пошаговая регрессия. Метод пошаговой регрессии начинается с построения простой регрессии. В анализ последовательно включают по одной объясняющей переменной. На каждом шаге проверяется значимость коэффициентов регрессии и оценивается

мультиколлинеарность переменных. Если оценка коэффициента получается незначимой, то переменная исключается и рассматривается другая объясняющая переменная. Таким образом пошагово определяются все составляющие регрессии с выполнением требования отсутствия мультиколлинеарности.

7) преобразование переменных. Например, пусть эмпирическое уравнение регрессии имеет вид: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$, причем x_1, x_2 коррелируют, можно определить регрессионные зависимости относительных величин:

$$\frac{\hat{y}}{x_1} = b_0 + b_1 \frac{x_2}{x_1} \text{ или } \frac{\hat{y}}{x_2} = b_0 + b_1 \frac{x_1}{x_2}.$$

Вполне вероятно, что в этих моделях мультиколлинераность будет отсутствовать.

8) Решением проблемы мультиколлинеарности может стать путь корректировки самого математического метода оценки параметров регрессионной модели. Так было предложено (Гоэрл) перед обращением матрицы $X'X$ прибавлять к ее диагональным элементам малые числа k. Это предупреждает вырождение матрицы, увеличивает ее определитель и уменьшает дисперсию параметров. Новые оценки (ридж-оценки) параметров вычисляются по формуле:

$$b = (X'X + kI)^{-1} X'Y.$$

От такой операции (увеличения диагональных элементов матрицы $X'X$ на малое число k) сумма квадратов ошибок увеличивается, но типичный характер зависимости остаточной дисперсии от k подобен графику на рис. 6.1 - до какой-то границы $k < k^*$ остаточная дисперсия возрастает незначительно.

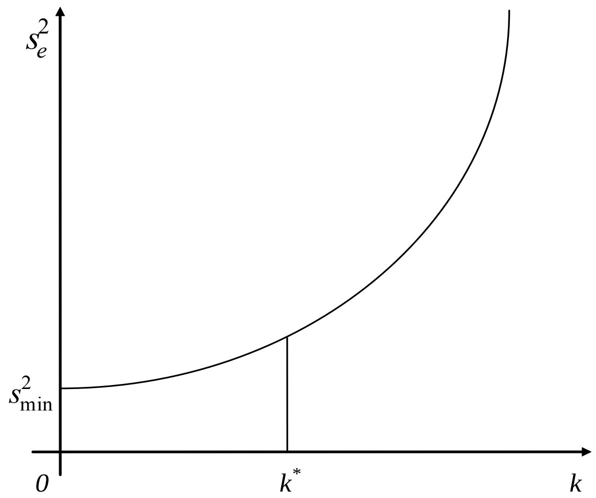


Рис. 6.1. Возрастание остаточной дисперсии s_e^2
при увеличении диагональных элементов матрицы $X'X$ на k

Устойчивость оценок параметров характеризуется суммой их дисперсий

$$L(k) = \sum_{i=1}^{m} s_{b_i}^2 .$$

Эту формулу можно преобразовать, если найти все собственные числа λ_i и собственные векторы U_i матрицы $X'X = UD_\lambda U'$. Здесь через D_λ обозначена диагональная матрица собственных чисел, а через U - ортогональная матрица собственных векторов. Вычисляем $U'b = g$. Теперь сумма дисперсий параметров модели может быть представлена в виде:

$$L(k) = \sum_{i=1}^{m} \frac{\lambda_i s_e^2(k) + k^2 g_i}{(\lambda_i + k)^2} .$$

При мультиколинеарности (вырождении матрицы $X'X$) одно или несколько собственных чисел λ_i близки к нулю и при $k = 0$ получается очень большое (если не бесконечное) значение $L(k)$. Введение даже очень малого k сразу исправляет ситуацию.

Существует несколько рекомендаций по выбору величины k; при компьютерной же реализации пошли по самому радикальному пути - вычисляется ряд оценок при различных значениях параметра $0 < k < 0{,}1$ и строятся графики «следа» зависимостей коэффициентов регрессии и остаточной дисперсии от этого параметра. Если имеется мультиколинеарность, то при $k \to 0$ начинается резкое возрастание некоторых (если не всех) коэффициентов регрессии. Выбираем такое наименьшее значение параметра k, при котором поведение всех коэффициентов регрессии уже стабилизировалось, но еще не начался интенсивный рост остаточной дисперсии.

Для сравнения различных альтернативных схем регрессионного анализа специалистами были предприняты многочисленные проверки на данных с различными особенностями. Первое место во всех таких проверках заняли ридж-оценки. МНК-оценки во всех проверках оказывались худшими по всем критериям качества. В связи с этим есть рекомендации: если мультиколинеарности нет, то ридж-оценки близки к МНК-оценкам, а в случае мультиколинеарности они значительно лучше, причем преимущество ридж-оценок увеличивается с ростом числа переменных m.

9) Метод главных компонент. Результатом применения метода главных компонент на исходной системе объясняющих переменных есть новые переменные, которые являются линейной комбинацией исходных. Исходную стандартизированную систему объясняющих переменных X_i преобразовывают некоррелируемой системой компонент F_i, имеющей свойства: первая компонента F_1 объясняет максимум общей изменчивости данных, вторая компонента F_2 - максимум остатка изменчивости и так далее. Несколько первых (главных) компонент объясняют почти всю изменчивость данных, а остальные компоненты отображают случайные ошибки и могут быть отброшены. В модели с главными компонентами выполняют обратный переход к исходным переменным X_i, тем самым обходят проблему мультиколлинеарности. Компоненты составляются как линейные комбинации:

$$F_i = u_1 X_1 + u_2 X_2 + ... + u_m X_m,$$

коэффициенты которой являются элементами собственных векторов $U_i = (u_1, u_2, ..., u_m)'$ корреляционной матрицы $R = (X'X)/n$; дисперсии компонент равны соответствующим собственным числам λ_i корреляционной матрицы.

Пример. Построить модель зависимости трудоемкости обработки детали (y) от количества операций (x_1), количества обработанных поверхностей (x_2), количества деталей (x_3) по данным, приведенным в табл. 6.1.

Таблица 6.1

Исходные и промежуточные данные

№	x_1	x_2	x_3	y	y_p	e
1	**2**	**3**	**4**	**5**	**6**	**7**
1	**239**	**449**	**52**	**72,4**	**63,54**	**8,86**
2	**240**	**450**	**50**	**72,7**	**60,33**	**12,37**
3	**251**	**465**	**53**	**73,4**	**63,29**	**10,11**
4	**256**	**470**	**65**	**75,5**	**73,27**	**2,23**
5	**258**	**475**	**67**	**78,5**	**80,48**	**-1,98**
6	**263**	**478**	**74**	**88,0**	**80,18**	**7,82**
7	**264**	**479**	**80**	**91,5**	**86,44**	**5,06**
8	**267**	**482**	**85**	**93,0**	**89,82**	**3,18**
9	**270**	**485**	**89**	**95,5**	**92,02**	**3,48**
10	**272**	**487**	**97**	**98,8**	**99,80**	**-1,00**
11	**274**	**492**	**99**	**105,8**	**107,01**	**-1,21**
12	**275**	**493**	**108**	**107,2**	**116,82**	**-9,62**

13	**276**	**494**	**111**	**107,9**	**119,53**	**-11,63**
14	**278**	**498**	**117**	**114,8**	**129,30**	**-14,50**
15	**283**	**502**	**128**	**122,7**	**135,91**	**-13,21**
16	**286**	**504**	**134**	**134,6**	**138,30**	**-3,70**
17	**290**	**509**	**137**	**147,5**	**140,64**	**6,86**
18	**291**	**512**	**144**	**150,3**	**152,45**	**-2,15**
19	**293**	**520**	**153**	**170,7**	**174,50**	**-3,80**
20	**295**	**525**	**168**	**192,1**	**197,11**	**-5,01**
21	**298**	**528**	**172**	**194,6**	**199,30**	**-4,70**
22	**300**	**532**	**173**	**198,8**	**203,15**	**-4,35**
23	**302**	**537**	**175**	**202,7**	**210,37**	**-7,67**
24	**304**	**539**	**178**	**206,1**	**212,23**	**-6,13**
25	**305**	**541**	**181**	**208,8**	**217,12**	**-8,32**
26	**308**	**546**	**185**	**216,6**	**223,67**	**-7,07**
27	**311**	**549**	**189**	**220,7**	**225,87**	**-5,17**
28	**318**	**554**	**193**	**229,0**	**220,32**	**8,68**
29	**319**	**557**	**205**	**239,0**	**238,05**	**0,95**
30	**322**	**560**	**215**	**254,6**	**247,35**	**7,25**
31	**325**	**564**	**224**	**267,9**	**257,64**	**10,26**
32	**327**	**569**	**231**	**275,0**	**270,78**	**4,22**
33	**328**	**573**	**238**	**275,4**	**284,77**	**-9,37**
34	**332**	**574**	**244**	**290,5**	**281,94**	**8,56**
35	**335**	**580**	**248**	**295,7**	**290,68**	**5,02**
36	**338**	**584**	**256**	**311,2**	**299,80**	**11,40**
37	**340**	**590**	**265**	**322,9**	**317,48**	**5,42**
38	**344**	**598**	**274**	**336,5**	**333,47**	**3,03**
39	**346**	**605**	**281**	**346,8**	**350,97**	**-4,17**
Средние	**295,46**	**524,33**	**157,38**	**181,68**	**181,68**	**0**
ДИСПР	**872,92**	**1803,4**	**4644,4**	**7163,32**	**7110,01**	**53,31**

Стандартным методом наименьших квадратов получено уравнение регрессии:

$$\hat{y} = -253{,}61 - 3{,}027x_1 + 2{,}181x_2 + 1{,}184x_3 \quad .$$

$$(t_b) \qquad\qquad (6{,}7) \qquad (6{,}0) \qquad (6{,}9)$$

Эта модель значимая в целом и имеет только значимые факторы (значения t-статистик Стьюдента $t_b > t_{0,05} = 2{,}0$). Коэффициент множественной корреляции очень высокий $R = 0{,}996$. В табл. 1 для каждого варианта данных приведены расчетные значения y_p и остатки модели $\varepsilon = y - y_p$. В последней строке табл.1 вычислены дисперсии всех переменных. Коэффициент детерминации можно вычислить двумя способами:

$$R^2 = 1 - \frac{s_\varepsilon^2}{s_y^2} = 1 - \frac{53{,}31}{7163{,}32} = 0{,}9926\,; \; RR = \frac{s_p^2}{s_y^2} = \frac{7110{,}01}{7163{,}32} = 0{,}9926\,.$$

Можно также проверить, что нет корреляции между остатками и расчетными значениями: $r_{\varepsilon y_p} = 0$ (согласно МНК, остатки ортогональны к расчетным значениям). На первый взгляд полученная модель полностью доброкачественная. Однако, имеем отрицательный коэффициент регрессии при x_1. Это интерпретируется так: чем больше операций при изготовлении детали, тем меньше трудоемкость? Построим эту модель зависимости в стандартизованных переменных:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

$$\hat{Y} = -1{,}057 X_1 + 1{,}094 X_2 + 0{,}953 X_3,$$

где β-коэффициенты оказались большими единицы (по модулю). Одной из причин такой ситуации может быть мультиколлинеарность. С табл. 6.2 следует, что между объясняющими переменными присутствуют очень большие корреляции. Фактически имеем не три объясняющие переменные, а три разные формы одной переменной.

Таблица 6.2

Корреляционная таблица

	x_1	x_2	x_3	y
x_1	**1**	**0,9967**	**0,9939**	**0,9813**
x_2	**0,9967**	**1**	**0,9955**	**0,9901**
x_3	**0,9939**	**0,9955**	**1**	**0,9925**
y	**0,9813**	**0,9901**	**0,9925**	**1**

Метод пошаговой регрессии в данном примере не устранил проблему подключения в модель тесно связанных групп показателей. Рассмотрим пошагово процесс составления регрессионной модели. Результативный признак y теснее связан с количеством деталей x_3. Учет одной переменной x_3 объясняет $(0{,}9925)^2 = 0{,}9851$ $(98{,}51\%)$ общей изменчивости результативного признака y. Трехфакторная модель объясняет 99,26% общей изменчивости y, то есть дополнительный вклад остальных факторов x_1, x_2 составляет только 0,75% изменчивости y. Учет x_3 фактически исчерпывает изменчивость всех переменных, от которых остаются только малые случайные остатки. Однако на следующем шаге в модель подключается x_1 (точнее малые остатки от этой переменной) и уже получается абсурдная модель: $\hat{y} = 254{,}03 - 1{,}174 x_1 + 1{,}738 x_3$. Далее подключается x_2. Следовательно, метод пошаговой регрессии не устраняет проблему мультиколлинеарности и в модель может войти группа тесно связанных показателей.

Используем метод гребневой (ридж-) регрессии, согласно которой к диагональным элементам корреляционной матрицы прибавляется малое число (ридж-параметр k). Учет этого малого параметра ведет к небольшой систематической ошибке в оценках коэффициентов регрессии, однако существенно снижает их случайную изменчивость. Оценки параметров стают несколько смещенными, но более эффективными. Известно, что дисперсии случайных ошибок коэффициентов регрессии пропорциональные диагональным элементам обратной матрицы $C = (r_{x_i x_j})^{-1}$. При мультиколлинеарности определитель корреляционной матрицы может оказаться близким к нулю (вырождение), тогда элементы обратной матрицы будут очень большими. Введение даже очень малого ридж-параметра препятствует чрезмерному возрастанию элементов обратной матрицы. В табл. 6.3 приведены обратные матрицы – исходная (C) и с учетом параметра $k = 0,05$ (C_k).

Таблица 6.3

Обратные матрицы C и C_k

$k = 0$			$k = 0,05$		
157,15398	**-127,646**	**-29,1128**	**12,279638**	**-6,29038**	**-5,659**
-127,6463	**215,8722**	**-88,0455**	**-6,290378**	**12,64669**	**-6,03662**
-29,1128	**-88,0455**	**117,5862**	**-5,659001**	**-6,03662**	**12,0323**

Таким образом, введение малого параметра $k = 0,05$ уменьшило элементы обратной матрицы более чем в 10 раз. При этом множитель пропорциональности
(MSE) изменился, но не существенно. При $k = 0,05$ модель имеет вид:

$$\hat{y} = -446,8 + 0,590x_1 + 0,713x_2 + 0,515x_3 \quad .$$

$$(s_b) \qquad\qquad (2,5) \qquad (4,8) \qquad (6,0)$$

Коэффициенты детерминации R^2 и RR теперь не равны:

$$R^2 = 0,9812 \text{ и } RR = 0,9473 .$$

Между остатками и расчетными значениями допущена небольшая корреляция:

$$r_{\varepsilon y_p} = 0,127 .$$

В отличие от исходной модели $k = 0$ здесь нет недоразумений. В табл. 6.4 приведены для разных k значения β-коэффициентов, а также R^2, RR и $r_{\varepsilon y_p}$.

На основании данных табл. 6.3 построены графики гребневого следа для каждой с названных характеристик.

Таблица 6.4

Значения коэффициентов регрессии в стандартизованных переменных

k	β_1	β_2	β_3	R^2	RR	r_{e,y_p}
0	**-1,057**	**1,094**	**0,953**	**0,993**	**0,993**	**0,000**
0,01	**-0,105**	**0,476**	**0,617**	**0,987**	**0,974**	**0,054**
0,02	**0,067**	**0,407**	**0,510**	**0,984**	**0,967**	**0,069**
0,03	**0,141**	**0,381**	**0,460**	**0,983**	**0,960**	**0,087**
0,04	**0,181**	**0,367**	**0,431**	**0,982**	**0,954**	**0,107**
0,05	**0,206**	**0,358**	**0,411**	**0,981**	**0,947**	**0,127**
0,06	**0,223**	**0,351**	**0,397**	**0,981**	**0,941**	**0,148**
0,07	**0,235**	**0,346**	**0,387**	**0,980**	**0,935**	**0,168**
0,08	**0,245**	**0,342**	**0,378**	**0,980**	**0,929**	**0,188**
0,09	**0,252**	**0,339**	**0,371**	**0,980**	**0,923**	**0,208**
0,1	**0,257**	**0,336**	**0,365**	**0,979**	**0,917**	**0,228**

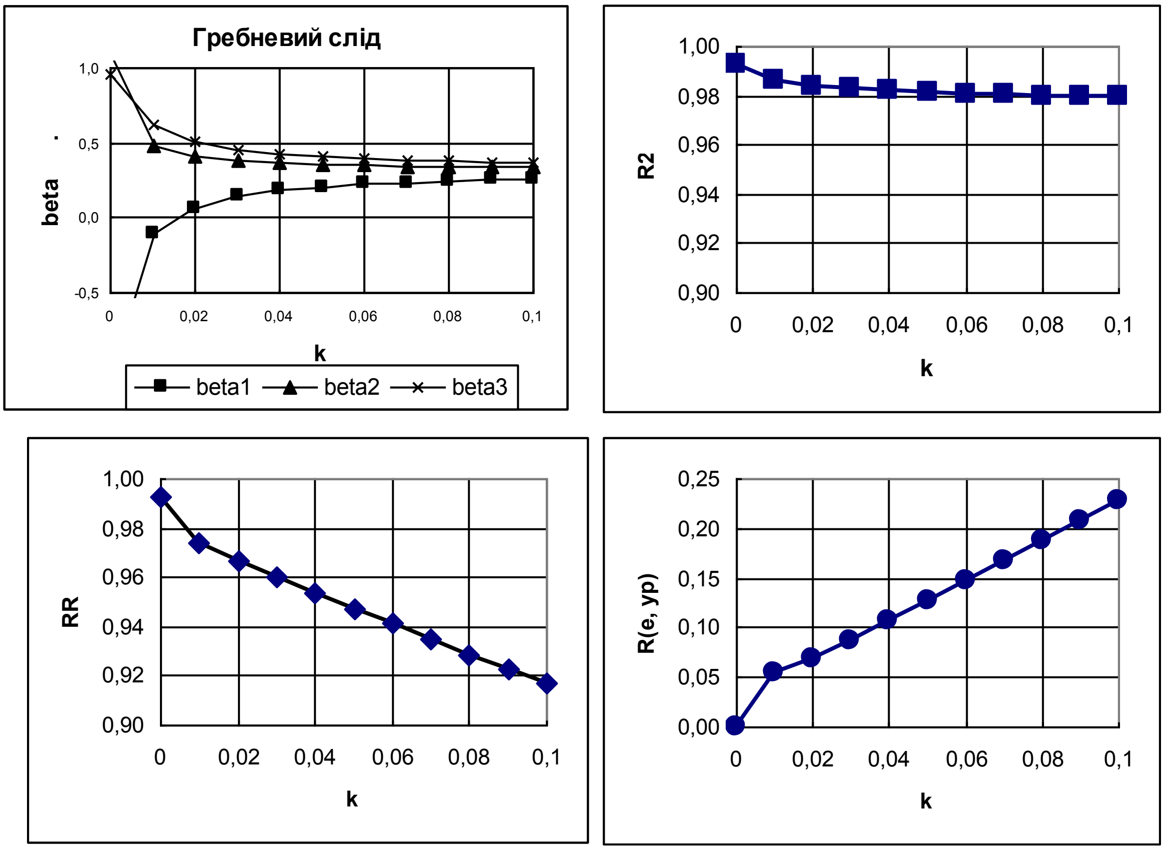


Рис. 2. Графики зависимости от ридж-параметра k

Видим, что оценка коэффициента детерминации по формуле $R^2 = 1 - \frac{s_\varepsilon^2}{s_y^2}$ с возрастанием ридж-параметра k уменьшается постепенно, не

так интенсивно, как оценка по формуле $RR=\frac{s_p^2}{s_y^2}$**. Поэтому как оценку**

дуется всегда использовать $R^2=1-\frac{s_\varepsilon^2}{s_y^2}$**. С возрастанием ридж-параметра** k **увеличивается корреляция расчетных значений** y_p **с остатками модели (это эффект внесения систематической ошибки). График на рис. 2 демонстрирует насколько неустойчивыми были начальные оценки и как существенно они изменились при совсем небольших значениях ридж-параметра. При** $k=0{,}02$ **ранее отрицательная оценка** β_1 **стала положительной, а при** $k=0{,}04$ **значения всех оценок стабилизировались. При этом коэффициент детерминации** R^2 **практически не изменился (он снизился от 0,996 к 0,982), между расчетными значениями и остатками появилась небольшая корреляция** $r_{\varepsilon y_p}=0{,}110$**. При** $k>5$ **все** β**-коэффициенты стают равны между собой – такой результат можно получить, используя метод главных компонент.**

Таблица 6.5

Собственные векторы и собственные числа корреляционной матрицы

№	U_1	U_2	U_3
1	**0,577325**	**-0,60081**	**0,552918**
2	**0,577633**	**-0,17808**	**-0,79664**
3	**0,577092**	**0,779302**	**0,24424**
λ	**2,990734**	**0,006256**	**0,00301**
%	**99,7%**	**0,2%**	**0,1%**
Σ	**99,7%**	**99,9%**	**100,0%**

С табл. 6.5 следует, что вклад одной первой компоненты составляет 99,7%, на вторую и третью – припадает 0,3%. Следовательно, достаточно учесть одну первую (главную) компоненту, которая является комбинацией исходных переменных с почти одинаковыми коэффициентами (элементы первого собственного вектора U_1):

$$F_1=0{,}5773X_1+0{,}5776X_2+0{,}5771X_3.$$

В конце получим уравнение регрессии в стандартизованных переменных:

$$\hat{Y}=0{,}3303X_1+0{,}3305X_2+0{,}3302X_3.$$

В однофакторной модели $\hat{Y}=0{,}9925X_3$ весь вклад трех равноправных факторов присвоено одной переменной.

В современных условиях развития программного обеспечения и компьютеров все перечисленные методы устранения мультиколлинеарности достаточно легко применимы. Для подтверждения правильности решения рекомендуется использовать несколько.

Вопросы для самопроверки:

1) Перечислить предпосылки МНК. Какие последствия их выполнимости или невыполнимости?

2) Какими свойствами должны обладать оценки полученные по МНК?

3) Объяснить суть проблемы мультиколлинеарности при построении множественной регрессионной модели.

4) Какие основные последствия мультиколлинеарности?

5) Как можно обнаружить мультиколлинеарность?

6) Перечислить основные методы устранения мультиколлинеарности. Объяснять их отличие.

Тема 7. Гетероскедастичность и методы ее определения. Обобщенный метод наименьших квадратов

7.1. Суть гетероскедастичности.

7.2. Последствия гетероскедастичности.

7.3. Методы определения и смягчения гетероскедастичности.

7.1. Суть гетероскедастичности

Одной из ключевых предпосылок МНК является условие постоянства дисперсий случайных отклонений – предпосылка дисперсия случайных отклонений ε_i постоянна $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$ для любых наблюдений i и j. Выполнимость данной предпосылки называется гомоскедастичность (постоянством дисперсии отклонений). Невыполнимость данной предпосылки называется гетероскедастичность (непостоянством дисперсий отклонений).

На практике гетероскедастичность не так уж и редка. Часто есть основание считать, что вероятностные распределения случайных отклонений ε_i при различных наблюдениях будут различными. Это не означает, что случайные отклонения обязательно будут большими при определенных наблюдениях и малыми – при других, но это означает, что априорная вероятность этого велика.

На рис. 7.1 приведены два примера линейной регрессии – зависимости потребления C от дохода I: $C = a + bI + \varepsilon$

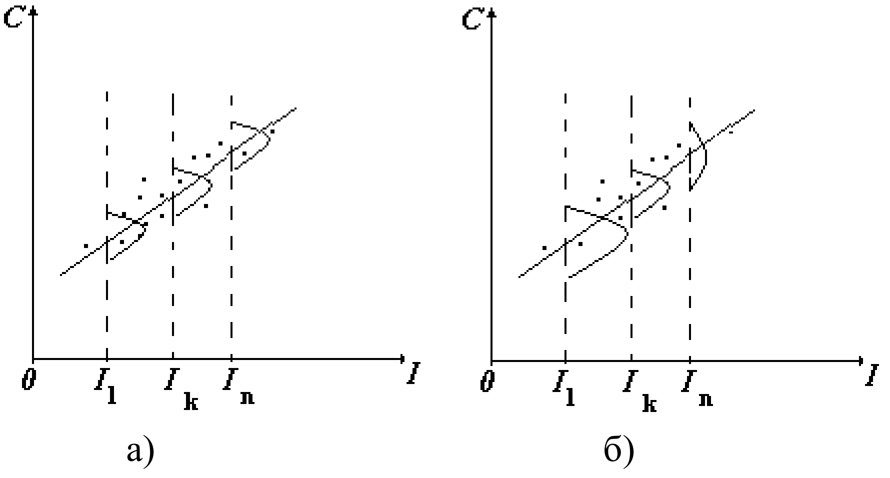


Рис. 7.1. а) гомоскедастичность остатков; б) гетероскедастичность остатков

В обоих случаях с ростом дохода растет среднее значение потребления. Но если на рис.1 а) дисперсия потребления остается одной и той же для различных уровней дохода, то на рис.1 б) при аналогичной зависимости среднего потребления от дохода дисперсия потребления не остается постоянной, а увеличивается с ростом дохода. Фактически это означает, что во втором случае субъекты с большим доходом в среднем потребляют больше, чем субъекты с меньшим доходом, и кроме того, разброс в их потреблении более существенен для большего уровня дохода. Люди с большим доходом имеют больший простор для его распределения. Реалистичность данной ситуации не вызывает сомнений. Проблема гетероскедастичности характерна для перекрестных данных (объекты-время) и довольно редко встречаются при рассмотрении временных рядов.

7.2. Последствия гетероскедастичности.

При невыполнимости данной предпосылки (при гетероскедастичности) последствия применения МНК будут следующими:

1. Оценки коэффициентов по-прежнему останутся несмещенными и линейными.

2. Оценки не будут эффективными. Увеличение дисперсии оценок снижает вероятность получения максимально точных оценок.

3. Дисперсии оценок будут рассчитываться со смещением. Смещенность появляется вследствие того, что не объясненная уравнением регрессии дисперсия

$S^2 = \dfrac{\sum \varepsilon_i^2}{n-m-1}$, которая используется при вычислении оценок дисперсий всех коэффициентов не является более несмещенной.

4. Вследствие все выводы, получаемые на основе соответствующих t- и F-статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получаемые при стандартных проверках качества оценок, могут быть ошибочными и приводить к неверным заключениям по построенной модели.

7.3. Методы определения и смягчения гетероскедастичности

В ряде случаев, зная характер данных, появление проблемы гетероскедастичности можно предвидеть и попытаться устранить этот недостаток еще на этапе спецификации. Однако значительно чаще эту проблему приходиться решать после построения уравнения регрессии. Сейчас не существует какого-либо однозначного метода определения гетероскедастичности. К настоящему времени для такой проверки разработано довольно большое число тестов и критериев для них. Наиболее популярными и наглядными есть: графический анализ отклонений, тест ранговой корреляции Спирмена, тест Парка, тест Глейзера, тест Голдфелда-Квандта.

Графический анализ остатков.

Использование графического представления отклонений позволяет определиться с наличием гетероскедастичности. В этом случае по оси абсцисс откладываются значения (x_i) объясняющей переменной X (либо линейной комбинации объясняющих переменных $Y = a + b_1x_1 + ... + b_mx_m$), а по оси ординат либо отклонения ε_i, либо их квадраты ε_i^2. Примеры таких графиков приведены на рис. 7.2.

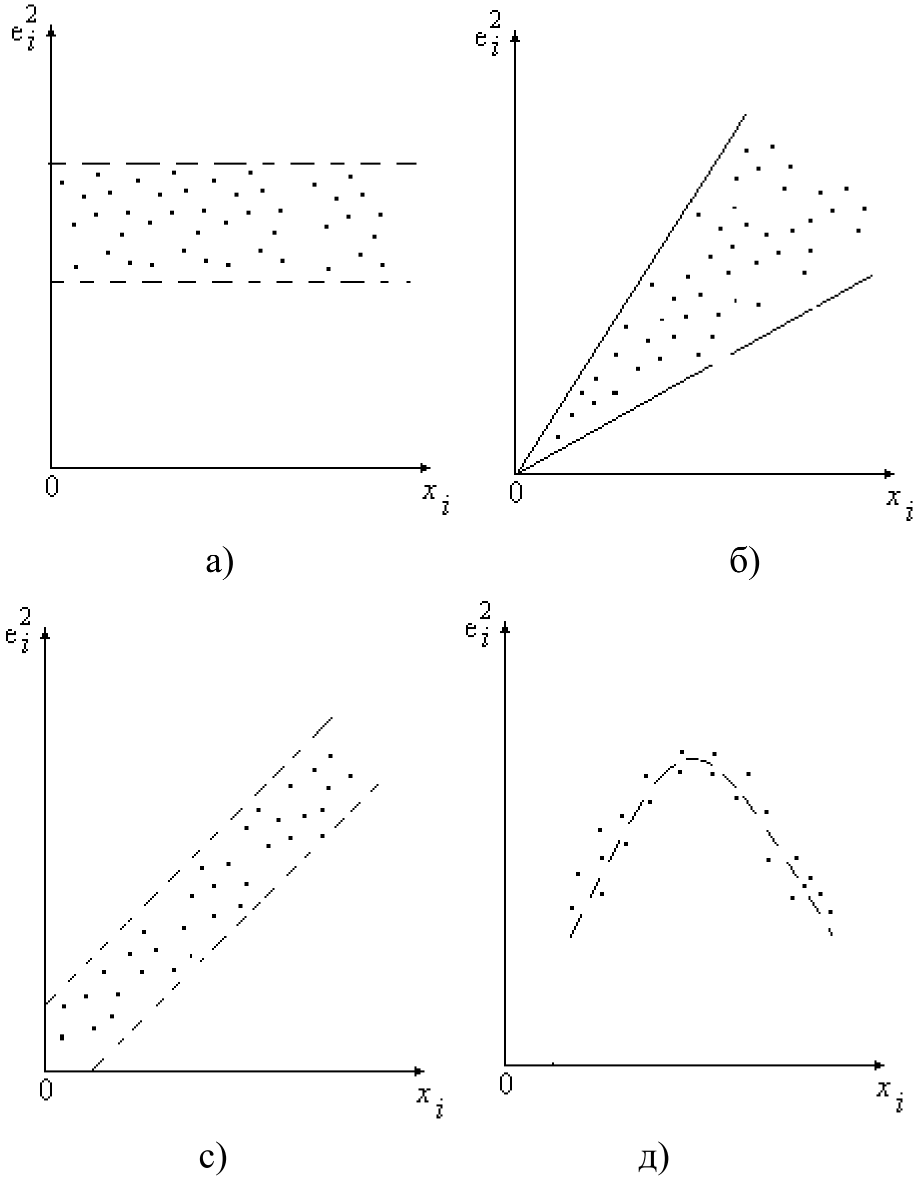


Рис. 7.2. Зависимости ε_i^2 от x_i

На рис. 7.2 а) все отклонения находятся внутри полосы постоянной ширины, параллельной оси абсцисс. Это говорит о независимости дисперсий ε_i^2 от значений переменной X и их постоянстве, т.е. в этом случае выполняются условия гомоскедастичности. Другие рисунки 7.2. б, с, д отражают большую вероятность наличия гетероскедастичности для рассматриваемых статистических данных.

Такой анализ наиболее целесообразен при большом количестве объясняющих переменных.

Тест ранговой корреляции Спирмена

При использовании данного теста предполагается, что дисперсия отклонений будет либо увеличиваться, либо уменьшаться с увеличением значений x. Поэтому для регрессии, построенной по МНК, абсолютные величины отклонений ε_i и значения x_i случайной величины X будут коррелированны. Значения x_i и ε_i ранжируются (упорядочиваются по величинам). Затем определяется коэффициент ранговой корреляции: $r_{x,\varepsilon} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2-1)}$, где d_i - разность между рангами x_i и ε_i. Доказано, что если коэффициент корреляции $\rho_{x,\varepsilon}$ для генеральной совокупности равен нулю, то статистика $t = \frac{r_{x,\varepsilon}\sqrt{n-2}}{\sqrt{1-r_{x,\varepsilon}^2}}$ имеет распределение Стьюдента с числом степеней свободы $\nu = n-2$. Следовательно, если наблюдаемое значение t-статистики превышает $t_{\text{кр}} = t_{\frac{\alpha}{2}, n-2}$, то необходимо отклонить гипотезу о равенстве нулю коэффициента корреляции $\rho_{x,\varepsilon}$, а, следовательно, и об отсутствии гетероскедастичности. В противном случае гипотеза об отсутствии гетероскедастичности принимается.

Если в модели регрессии больше чем одна объясняющая переменная, то проверка гипотезы может осуществляться с помощью t-статистики для каждой из них отдельно.

Пример. *Продолжение пример лекции 2.* Проверить гипотезу об отсутствии гетероскедастичности с помощью теста ранговой Спирмена. Доверительная вероятность $p = 95\%$.

x_i	ε_i	$\lvert\varepsilon_i\rvert$	d_1	d_2	$d = d_1 - d_2$	d^2
2	0	0	5	5	0	0
3	-0,09	0,09	4	2	2	4
4	0,12	0,12	3	1	2	4
5	0,03	0,03	2	4	-2	4
6	-0,06	0,06	1	3	-2	4
Сумма						16

$n = 5$, $r_{x,\varepsilon} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2-1)} = 1 - 6 \cdot \frac{16}{5(5^2-1)} = 0,2$.

$\alpha = \frac{1-p}{2} = \frac{1-0,95}{2} = 0,25$. По t-таблицам находим граничную точку $t_{\text{кр}} = t_{\frac{\alpha}{2}, n-2} = t_{0,025;\, 5-2} = 3,182$.

Статистика $t = \frac{r_{x,\varepsilon}\sqrt{n-2}}{\sqrt{1-r_{x,\varepsilon}^2}} = \frac{0{,}2\sqrt{5-2}}{\sqrt{1-0{,}2^2}} = 0{,}352 < 3{,}182$. Мы принимаем гипотезу об отсутствии гетероскедастичности на уровне значимости 5%.

Тест Парка

Критерий Парка предполагает следующие этапы:

1) Строится уравнение регрессии $y_i = a + b_1 x_i + \varepsilon_i$.
2) Для каждого наблюдения определяется $\ln \varepsilon_i^2 = \ln(y_i - \hat{y}_i)^2$.
3) Строится регрессия $\ln \varepsilon_i^2 = \alpha + \beta \ln x_i + v_i$, где $\alpha = \ln \sigma^2$. Для множественной регрессии строится эта зависимость для каждой объясняющей переменной.
4) Проверяется статистическая значимость коэффициента β на основе t-статистики $t = \frac{\beta}{\sigma_\beta}$. Если коэффициент β статистически значим, то это означает наличие связи между $\ln \varepsilon_i^2$ и $\ln x_i$, т.е. гетероскедастичности в статистических данных.

Применение теста Парка в некоторых случаях может привести к необоснованным выводам, поэтому он дополняется другими тестами.

Тест Голдфелда-Квандта

В тесте Голдфелда-Квандта предполагается, что стандартное отклонение σ_ε пропорционально значению переменной x в этом наблюдении: $\sigma_{\varepsilon_i}^2 = \sigma^2 x_i^2 \quad (i = \overline{1,n})$. Также предполагается, что ε_i имеет нормальное распределение и отсутствует автокорреляция и все n наблюдений упорядочиваются по величине x. Эта упорядоченная выборка делится на три примерно равные части $k, n-2k, k$ соответственно. Для каждой из выборок объема k оценивается свое уравнение регрессии и находятся суммы квадратов отклонений $S_1 = \sum_{i=1}^{k} \varepsilon_i^2$ и $S_3 = \sum_{i=n-k+1}^{k} \varepsilon_i^2$ соответственно.

При заданной доверительной вероятности $p, \alpha = 1 - p$ по F-таблицам находим граничную точку $F_{\alpha,\, k-m-1,\, k-m-1}$, где m-число факторов модели;

рассчитываем значение $F = \frac{S_3}{S_1}$. Если $F < F_{\alpha,\, k-m-1,\, k-m-1}$, то на уровне значимости α принимается гипотеза об отсутствии гетероскедастичности. В противном случае гипотеза об отсутствии гетероскедастичности отклоняется. Для множественной регрессии тест обычно проводится для того фактора, который в максимальной степени связан с σ_{ε_i}. При этом выбирают $k > m+1$. Для убедительности данный тест можно рассчитать для каждого фактора.

Пример. Имеется регрессионная линейная модель с $m = 2$ фактора, $n = 30$ наблюдений. Для первых и последних $k = 11$ наблюдений суммы квадратов отклонений $S_1 = 20$ и $S_3 = 45$ соответственно. С помощью теста Голдфелда-Квандта проверить гипотезу об отсутствии гетероскедастичности.

Пусть доверительная вероятность $p = 95\%$. Тогда $\alpha = 0{,}05$. По F-таблицам находим граничную точку $F_{0{,}05,\,8,\,8} = 3{,}44$. Рассчитываем $F = \frac{S_3}{S_1} = \frac{45}{20} = 2{,}25 < 3{,}44$. На уровне значимости 5% принимается гипотеза от отсутствии гетероскедастичности.

Метод взвешенных наименьших квадратов

При установлении гетероскедастичности возникает необходимость преобразования модели с целью устранения данной проблемы. Вид преобразования зависит от того, известны или нет дисперсии σ_ε^2 отклонений ε_i. Если известны σ_ε^2, то рекомендуется применять метод взвешенных наименьших квадратов (ВНК). При этом следует разделить каждое наблюдение на соответствующее ему значение дисперсии. Например, пусть имеем $y_i = a + b_1 x_i + \varepsilon_i$. Разделим обе части уравнения на известное $\sigma_{\varepsilon_i} = \sqrt{\sigma_{\varepsilon_i}^2}$:

$$\frac{y_i}{\sigma_{\varepsilon_i}} = \frac{a}{\sigma_{\varepsilon_i}} + b_1 \frac{x_i}{\sigma_{\varepsilon_i}} + \frac{\varepsilon_i}{\sigma_{\varepsilon_i}}.$$

Пусть $\frac{y_i}{\sigma_{\varepsilon_i}} = y_i^*$, $\frac{x_i}{\sigma_{\varepsilon_i}} = x_i^*$, $\frac{\varepsilon_i}{\sigma_{\varepsilon_i}} = v_i$, $\frac{1}{\sigma_{\varepsilon_i}} = z_i$. Получили уравнение регрессии без свободного члена, но с дополнительной объясняющей переменной Z и с преобразованным отклонением v_i:

$$y_i^* = az_i + b_1 x_i^* + v_i .$$

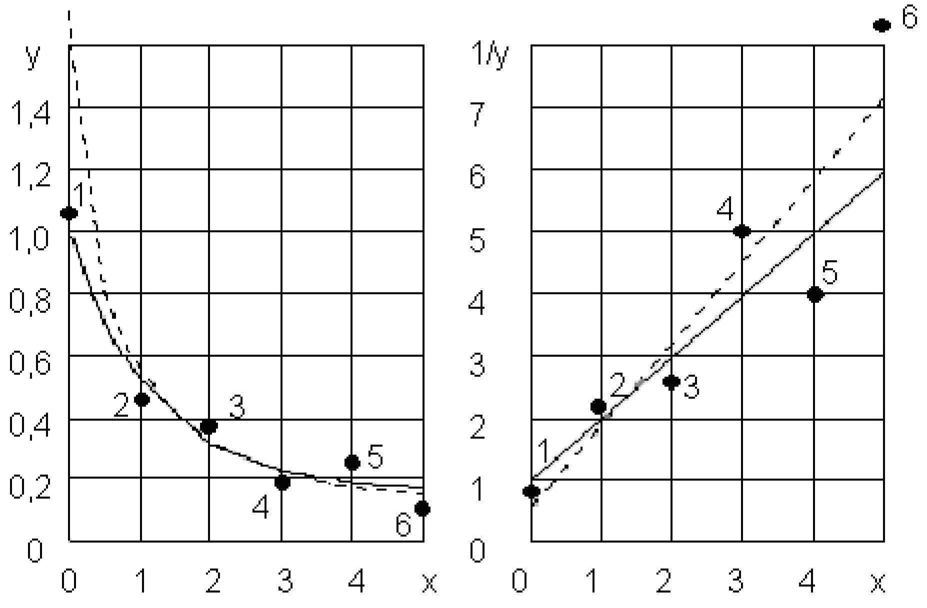
При этом для всех v_i выполняется условие гомоскедастичности.

Пример. Пусть значения y порождаются точной зависимостью $\eta = \dfrac{1}{1+x}$, которым присвоены погрешности $\varepsilon = \pm 0{,}05$, т.е. $y = \eta(x) + \varepsilon$; аргумент x приобретает последовательные целые значения 0, 1, 2, 3, 4, 5, 6. Предполагается, что форма связи известна: $y_p = \dfrac{1}{a + b_1 x}$.

После функционального преобразования $Y = \dfrac{1}{y}$ модель формально становится линейной $Y_p = a + b_1 x$ и ее параметры могут быть оценены стандартной процедурой МНК. В табл. 11.1 и на рис. 3 приведены начальные данные $(x, \eta, \varepsilon, y)$, преобразованные значения зависимой переменной $\left(\dfrac{1}{\eta}, \dfrac{1}{y}\right)$ и новая система ошибок $\left(e = \dfrac{1}{\eta} - \dfrac{1}{y}\right)$.

n	x	η	ε	y	$\dfrac{1}{\eta}$	$\dfrac{1}{y}$	e
1	0	1	0,05	1,050	1	0,952	-0,048
2	1	0,5	-0,05	0,450	2	2,222	0,222
3	2	0,333	0,05	0,383	3	2,611	-0,389
4	3	0,25	-0,05	0,200	4	5	1
5	4	0,2	0,05	0,250	5	4	-1
6	5	0,167	-0,05	0,117	6	8,547	2,546

С таблицы видно, что увеличение разброса преобразованных данных $\dfrac{1}{y}$ при увеличении переменной x. На рис. 7.3 б) это четко видно, причем последняя точка $\left(x, \dfrac{1}{y}\right)$ вышла за поле графика и похожа на выброс. Этот выброс перетянул в свою сторону всю линию регрессии.



а) *б*)

Рис. 7.3. Нелинейная зависимость в исходных переменных (а) и после функциональных преобразований (б), где • – эмпирические точки; — порождающая зависимость; --- восстановлена по МНК

Имеем следующие выкладки для вычисления параметров модели $\frac{1}{y} = a + b_1 x + e$. Система нормальных уравнений имеет вид:

$$\left[\frac{1}{y}\right] = b_0 + b_1 [x];$$

$$\left[\frac{x}{y}\right] = b_0 [x] + b_1 [x^2].$$

Эти уравнения реализуют условия ортогональности ошибок для каждого члена модели $[e] = 0$, $[ex] = 0$. Все необходимые расчеты выполнены в таблице.

n	x	x^2	y	$\frac{1}{y}$	$\frac{x}{y}$	$\frac{1}{y_p}$	y_p
1	0	0	1,050	0,952	0	0,624	1,601
2	1	1	0,450	2,222	2,222	1,930	0,518
3	2	4	0,383	2,611	5,222	3,236	0,309
4	3	9	0,200	5,000	15,000	4,541	0,220
5	4	16	0,250	4,000	16,000	5,847	0,171
6	5	25	1,117	8,547	42,735	7,153	0,140
Суммы	15	55		23,332	81,179		

Получим следующую систему нормальных уравнений:

$$23{,}332 = a \cdot 6 + b_1 \cdot 15;$$

$$81{,}179 = a \cdot 15 + b_1 \cdot 55.$$

Решив систему уравнений, получим такие значения параметров модели $a = 0{,}6245$, $b_1 = 1{,}3057$, они существенно отличаются от параметров зависимости $\eta = \dfrac{1}{1+x}$, где $a = b_1 = 1$. Вычисленные значения

$$\frac{1}{y_p} = 0{,}6245 + 1{,}3057x;$$

$$y_p = \frac{1}{0{,}6245 + 1{,}3057x};$$

приведены в последних двух столбцах таблицы, графики линий регрессии изображены пунктирными линиями на рис. 4. При $x = 0$ вместо ожидаемого значения $\eta = 1$ (с ошибкой близкой к $\varepsilon = 0{,}05$) получим $y_p = 1{,}6$, что является неправильным. Вид графика 7.3 (б) и его расположение относительно эмпирических точек вызывает сомнение о правильности выбора самой формы связи.

Следовательно, любые нарушения начальных гипотез приводят к ошибкам в оценивании параметров модели, причем ошибки могут быть исправлены путем модернизации стандартной процедуры оценивания. Поэтому для устранения гетероскедастичности в данном примере умножим обе части уравнения

$$\frac{1}{y} = a + b_1 x + e$$

на y^2 и получим ($\varepsilon = ey^2$):

$$y = ay^2 + b_1 xy^2 + \varepsilon.$$

Для этой модели система нормальных уравнений имеет вид:

$$\sum y^3 = a\sum y^4 + b_1 \sum xy^4;$$

$$\sum xy^3 = a\sum xy^4 + b_1 \sum x^2 y^4;$$

(условия ортогональности ошибок к каждому члену модели $\sum \varepsilon y^2 = 0$, $\sum \varepsilon xy^2 = 0$). Необходимые суммы подсчитаны в таблице:

n	x	y	y^3	y^4	xy^3	xy^4	x^2y^4	$\frac{1}{y_p}$	y_p	η
1	0	1,050	1,158	1,216	0	0	0	0,959	1,043	1
2	1	0,450	0,091	0,041	0,091	0,041	0,041	1,901	0,526	0,5
3	2	0,383	0,056	0,026	0,112	0,043	0,086	2,843	0,352	0,333
4	3	0,200	0,008	0,002	0,024	0,480	0,014	3,786	0,264	0,25
5	4	0,250	0,016	0,004	0,062	0,016	0,062	4,728	0,212	0,2
6	5	1,117	0,002	0,000	0,008	0,001	0,005	5,670	0,176	0,167
Суммы			1,330	1,284	0,298	0,105	0,209			

Откуда получим такую систему нормальных уравнений:

$$1{,}330 = a \cdot 1{,}284 + b_1 \cdot 0{,}105\,;$$

$$0{,}298 = a \cdot 0{,}105 + b_1 \cdot 0{,}209\,.$$

Решением новой системы нормальных уравнений являются значения параметров $a = 0{,}959,\ \ b_1 = 0{,}942$, которые близки к ожидаемым $a = \beta_1 = 1$ (параметры модели $\eta = \dfrac{1}{1+x}$). В последних колонках таблицы вычислены значения $\dfrac{1}{y_p}$ и y_p. Для сравнения приведены ординаты η. Соответствие между заданной (η) и воспроизведенной методом наименьших квадратов зависимостями очень большое. Таким образом, замена стандартной процедуры МНК на обобщенную процедуру взвешенных наименьших квадратов обеспечивает выполнение второй гипотезы Гаусса-Маркова о равноточности наблюдений.

На практике рекомендуют применять несколько методов определения гетероскедастичности и способов ее корректировки.

Вопросы для самопроверки:

1) Объяснить суть гетероскедастичности.
2) Какие известны последствия гетероскедастичности?
3) Перечислить методы определения гетероскедастичности.
4) В чем суть теста ранговой корреляции Спирмена?
5) В чем суть теста Парка?
6) В чем суть теста Голдфелда-Квандта?

7) Перечислить методы смягчения гетероскедастичности.
8) В чем суть метода взвешенных наименьших квадратов?
9) Почему при наличии гетероскедастичности ВНК позволяет получить более эффективные оценки, чем обычный МНК?

Тема 8. Автокорреляция остатков модели и методы ее устранения

8.1. Суть и причины автокорреляции

8.2. Последствия автокорреляции. Методы определения автокорреляции

8.3. Методы устранения автокорреляции

8.1. Суть и причины автокорреляции

В динамических рядах важную роль играет фактор времени, который вмещая в себя многие другие факторы развития, вызывает направление изменение социально-экономических явлений. Члены одного и того динамического ряда связаны между собой: предыдущие члены влияют на последующие. Это явление называется автокорреляцией. Поэтому, прежде чем находить количественную оценку связи между временными рядами, необходимо проверить существование автокорреляции. Различают автокорреляцию переменных y и x_i и автокорреляцию остатков модели.

Важной предпосылкой построения качественной регрессионной модели по МНК является независимость значений случайных отклонений ε_i от значений отклонений во всех других наблюдениях. Отсутствие зависимости гарантирует отсутствие коррелированности между любыми отклонениями $\sigma(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$ и, в частности, между соседними отклонениями $\sigma(\varepsilon_{i-1}, \varepsilon_i) = 0$.

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные данные). Автокорреляция остатков (отклонений) обычно встречается в регрессионном анализе при использовании данных временных рядов. Чаще всего встречается так называемая положительная автокорреляция, по сравнению отрицательная.

Среди основных причин, вызывающих появление автокорреляции, выделяют ошибки спецификации, инерцию в изменении экономических показателей, эффект паутины, сглаживание данных. Неучет в модели важной переменной или неправильный выбор формы зависимости приводит к системным отклонениям точек наблюдений от линий регерссии, что может обусловить автокорреляцию. Многие экономические показатели (инфляция, безработица, ВНП и другие) обладают определенной цикличностью, связанной с волнообразностью деловой активности. Во многих производственных и других сферах экономические показатели реагируют на изменение экономических условий с запаздыванием (временным лагом). А также, данные по некоторому продолжительному временному периоду получают усреднением данных по составляющим его подинтервалам. Это может привести к определенному сглаживанию колебаний, которые имелись внутри рассматриваемого периода, что в свою очередь может послужить причиной автокорреляции. Одной из причин возникновения автокорреляции остатков может быть, что применяемый критерий не может служить объективным показателем автокрреляции.

8.2. Последствия автокорреляции. Методы определения автокорреляции

Последствия автокорреляции в определенной степени такие же как и последствия гетероскедастичности. Дисперсии оценок будут смещенными. Часто дисперсии, которые рассчитываются по стандартным формулам, являются заниженными. Это приводит к увеличению t-статистик. Что, в свою очередь, влечет признание значимыми объясняющие переменные, которые в действительности незначимы. А также, выводы по t-статистике и F-статистике могут быть не верными, что ухудшает статистическое качество модели.

Существует несколько методов определения автокорреляции.

Графический метод. Рассмотрим графический метод, увязывающий отклонения ε_i с моментами t их получения (их порядковыми номерами i). Это так называемые последовательно-временные графики.

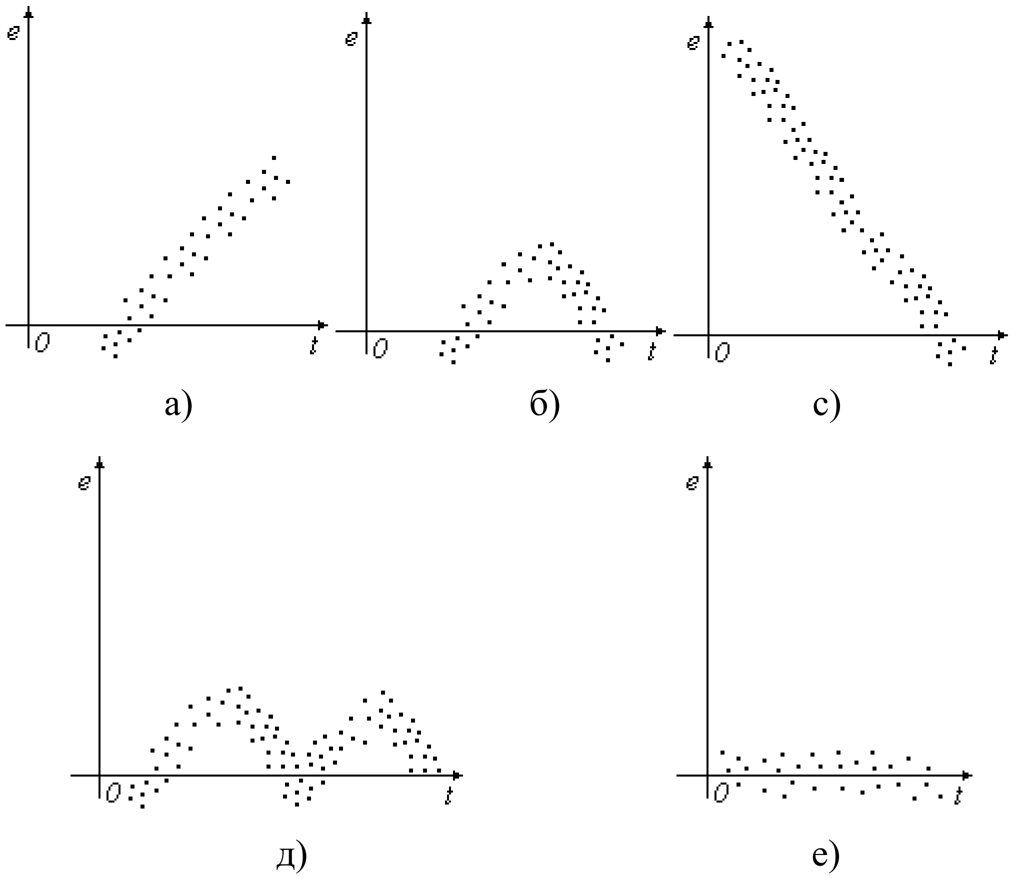


Рис. 8.1

На рис. 8.1 а – д автокорреляция имеет место, а рис. 8.1 е) свидетельствует об отсутствии автокорреляции. В современных компьютерных прикладных программах для решения задач по эконометрике аналитическое выражение регрессии дополняется графическим представлением результатов.

Критерий Дарбина-Уотсона. Наиболее известным критерием обнаружения автокорреляции первого порядка является критерий Дарбина-Уотсона. На основе вычисленной статистики DW Дарбина-Уотсона делается вывод об автокорреляции:

$$DW = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T} \varepsilon_t^2}. \qquad (1)$$

Таким образом статистика Дарбина-Уотсона тесно связана с выборочным коэффициентом корреляции $r_{\varepsilon_t \varepsilon_{t-1}}$: $DW \approx 2(1 - r_{\varepsilon_t \varepsilon_{t-1}})$. Таким образом, $0 \le DW \le 4$ и ее значения могут указать на наличие либо отсутствие автокорреляции. Если $r_{\varepsilon_t \varepsilon_{t-1}} = 0$ (автокорреляция отсутствует), то $DW \approx 2$. Если $r_{\varepsilon_t \varepsilon_{t-1}} = 1$ (положительная автокорреляция), то $DW \approx 0$. Если $r_{\varepsilon_t \varepsilon_{t-1}} = -1$

(отрицательная автокорреляция), то $DW \approx 4$. В критерии Дарбина-Уотсона при заданном уровне значимости α, числа наблюдений n и количества объясняющих переменных m определяются два значения: d_1 - нижняя граница и d_u - верхняя граница. Общая схема критерия Дарбина-Уотсона следующая:

1. Определяются значения отклонений $\varepsilon_t = y_t - \hat{y}_t$.
2. По формуле (1) рассчитывается статистика DW.
3. По таблице критических точек Дарбина-Уотсона определяются два числа d_l и d_u, затем делаются выводы по правилу:

$0 \le DW \le d_l$ - существует положительная корреляция;

$d_l \le DW \le d_u$ - вывод о наличии автокорреляции не определен;

$d_u \le DW \le 4 - d_u$ - автокорреляция отсутствует;

$4 - d_u \le DW \le 4 - d_l$ - вывод о наличии автокорреляции не определен;

$4 - d_l \le DW \le 4$ - существует отрицательная автокорреляция.

8.3. Методы устранения автокорреляции

Основной причиной наличия случайного члена в модели являются несовершенные знания о причинах и взаимосвязях, определяющих то или иное значение зависимой переменной. Поэтому свойства случайных отклонений, в том числе и автокорреляция, в первую очередь зависит от выбора формулы зависимости и состава объясняющих переменных. Так как автокорреляция чаще всего вызывается неправильной спецификацией модели, то необходимо, прежде всего, скорректировать саму модель. Также можно попробовать изменить форму зависимости (например, линейную на лог-линейную или гиперболическую). Можно воспользоваться авторегрессионным преобразованием: положив $y_1^* = y_t - \rho\, y_{t-1}$, $x_t^* = x_t - \rho\, x_{t-1}$, $\beta_0^* = \beta_0 (1 - \rho)$ получим $y_t^* = \beta_0^* + \beta_1 x_t^* + \nu_t$. При таком преобразовании при малых выборках может привести к потере эффективности. Эта проблема обычно преодолевается с помощью поправки Прайса-Винстена: $x_1^* = \sqrt{1-\rho^2} \cdot x_1$, $y_1^* = \sqrt{1-\rho^2} \cdot y_1$.

Определение ρ на основе статистики Дарбина-Уотсона .

Статистики Дарбина-Уотсона тесно связана с коэффициентом корреляции между соседними отклонениями через соотношение

$DW \approx 2(1 - r_{\varepsilon_t \varepsilon_{t-1}})$. Тогда в качестве оценки коэффициента ρ может быть взят коэффициент $r = r_{\varepsilon_t \varepsilon_{t-1}}$. Тогда имеем

$r \approx 1 - \dfrac{DW}{2}$. Этот метод оценивания неплох при большом числе наблюдений. В этом случае оценка r параметра ρ будет достаточно точной.

Существуют другие методы определения ρ**,** например, на основе статистики Дарбина-Уотсона, метода Кохрана-Оркатта, метод Хилдрета-Лу, метода первых разностей.

Итерационный метод Кохрана-Оркатта

Когда автокорреляцией пренебречь нельзя, в методе рекомендуется учесть ее в модели:

$$y_t = a + b_1 x_{1,t} + b_1 x_{2,t} + e_t,$$

$$y_{t-1} = a + b_1 x_{1,t-1} + b_1 x_{2,t-1} + e_{t-1},$$

$$e_t = \rho e_{t-1} + \varepsilon_t,$$

где ε_t - некоррелированные ошибки.

Преобразуем модель к виду с некоррелированными ошибками:

$$v_t = a u_0 + b_1 u_{1,t} + b_1 u_{2,t} + \varepsilon_t,$$

где $u_0 = 1 - \rho,\ u_{1,t} = x_{1,t} - \rho x_{1,t-1},\ u_{2,t} = x_{2,t} - \rho x_{2,t-1}, v_t = y_t - \rho y_{t-1}$.

Предлагается выполнять расчеты последовательным приближением. Сначала получают начальные значения параметров без учета автокорреляций, вычисляют остатки модели e_t и оценку коэффициента автокорреляции ρ. Дальше составляют новую модель с некоррелированными ошибками:

$$y_t - \rho y_{t-1} = a(1 - \rho) + b_1(x_{1,t} - \rho x_{1,t-1}) + b_2(x_{2,t} - \rho x_{2,t-1}),$$

где значения ρ считается известным как оценка на предыдущей итерации.

На последующих итерациях находят уточненные МНК-оценки параметров регрессии и новые оценки для ρ. Ожидается, что новая оценка коэффициента регрессии может существенно отличаться от начальной оценки, полученной при $\rho = 0$.

Таким образом, при установлении автокорреляции необходимо прежде всего проанализировать правильность спецификации модели. Если после действий усовершенствования регрессии автокорреляция по-прежнему имеется, то считают, что это возможно связано с внутренними свойствами ряда отклонений ε_t. Рекомендуется выполнить преобразования, устраняющие автокорреляцию. А именно: авторегрессионная схема первого порядка $AR(1)$, при этом не обходимо оценить коэффициент корреляции между отклонениями. Это может быть сделано различными методами: на основе статистики Дарбина-Уотсона, Кохрана-Оркатта, Хилдрета-Лу и других методов. Если среди объясняющих переменных есть лаговая зависимая переменная наличие автокорреляции устанавливается с помощью h-статистики Дарбина. Для ее устранения рекомендуется импользовать метод Хилдрета-Лу.

Пример. Пусть объем предложения товара y зависит от цены товара x_1 и зарплаты сотрудников x_2: $\hat{y} = 90{,}74 + 0{,}88x_1 - 7{,}32x_2$ (табл. 8.1). Определить наличие автокорреляции с помощью критерия Дарбина-Уотсона.

Таблица 8.1

y_i	x_1	x_2	$\hat{y}_i$	ε_i	ε_i^2	$\varepsilon_i - \varepsilon_{t-1}$	$(\varepsilon_i - \varepsilon_{t-1})^2$
20	10	12	11,7	8,3	68,89		
35	15	10	30,74	4,26	18,15	-4,04	16,32
30	20	9	42,46	-12,46	155,25	-16,72	279,56
45	25	9	46,86	-1,86	3,46	10,6	112,36
60	40	8	67,38	-7,38	54,46	-5,52	30,47
70	37	8	64,74	5,26	27,67	12,64	159,77
75	43	6	84,66	-9,66	93,32	-14,92	22,61
90	35	4	92,26	-2,26	5,11	7,4	54,76
105	40	4	96,66	8,34	69,56	10,6	112,36
110	55	5	102,54	7,46	55,65	-0,88	0,77
640	**320**	**75**			**551,52**		**988,98**

$$DW = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T} \varepsilon_t^2} = \frac{988{,}98}{551{,}52} = 1{,}793.$$

По таблице распределения Дарбина-Уотсона находим d_l=0,697 и d_u=1,641. Тогда

$4 - d_u$=4−1,641=2,359. Поскольку $d_u \le DW \le 4 - d_u$ (1,641<1,793<2,359), то гипотеза об отсутствии автокорреляции остатков не отклоняется на уровне значимости 0,05. Это является одним из подтверждений высокого качества модели.

Пример. В табл. 8.2 представлены сведения в течении 24 лет о затратах на жилье в США (C, млрд дол) в зависимости от прибыли (W, млрд дол) и индекса реальных цен на жилье (P, для 1972 г. принято P = 100) [Доугерти].

Методом наименьших квадратов в логарифмическом масштабе получено следующее уравнение с коэффициентов детерминации $R^2 = 0{,}9908$:

$$\ln C = -1{,}51 + 1{,}18 \ln W - 0{,}35 \ln p + e,$$

$$(t) \quad (0{,}8) \quad (21{,}0) \quad (1{,}1)$$

Таблица 8.2

Затраты на жилье (C) в зависимости от прибыли (W) и индекса цен (P)

Рік	***C***	***W***	***P***	***Y = lnC***	***Const***	***X_1 = lnW***	***X_2 = lnP***	***Y_p = lnCp***	***e***
1960	64,0	489,7	104,5	4,159	1	6,194	4,649	4,140	0,019
1961	67,0	503,8	105,1	4,205	1	6,222	4,655	4,171	0,033
1962	70,7	524,9	105,0	4,258	1	6,263	4,654	4,220	0,039
1963	74,0	542,3	104,8	4,304	1	6,296	4,652	4,259	0,045
1964	77,4	580,8	104,5	4,349	1	6,364	4,649	4,341	0,008
1965	81,6	616,3	104,0	4,402	1	6,424	4,644	4,412	-0,010
1966	85,3	646,8	102,6	4,446	1	6,472	4,631	4,474	-0,027
1967	89,1	673,5	102,2	4,490	1	6,512	4,627	4,523	-0,033
1968	93,5	701,3	100,9	4,538	1	6,553	4,614	4,575	-0,037
1969	98,4	722,5	100,0	4,589	1	6,583	4,605	4,613	-0,024
1970	102,0	751,6	99,6	4,625	1	6,622	4,601	4,661	-0,036
1971	106,4	779,2	100,0	4,667	1	6,658	4,605	4,702	-0,035
1972	112,5	810,3	100,0	4,723	1	6,697	4,605	4,748	-0,025
1973	118,2	865,3	99,1	4,772	1	6,763	4,596	4,828	-0,056
1974	124,2	858,4	95,1	4,822	1	6,755	4,555	4,833	-0,011
1975	128,3	875,8	93,3	4,854	1	6,775	4,536	4,864	-0,009
1976	134,9	906,8	93,7	4,905	1	6,810	4,540	4,903	0,002
1977	141,3	942,9	94,5	4,951	1	6,849	4,549	4,946	0,005
1978	148,5	988,8	94,7	5,001	1	6,896	4,551	5,001	-0,001
1979	154,8	1015,5	93,8	5,042	1	6,923	4,541	5,036	0,006
1980	159,8	1021,6	93,0	5,074	1	6,929	4,533	5,046	0,028
1981	164,8	1049,3	94,2	5,105	1	6,956	4,545	5,073	0,032
1982	167,5	1058,3	96,7	5,121	1	6,964	4,572	5,074	0,047
1983	171,3	1095,4	99,2	5,143	1	6,999	4,597	5,105	0,038

В скобках приведены значения статистик Стьюдента для каждого коэффициента регрессии. Незначимый по критерию Стьюдента член - $0{,}35 \ln P$

не следует убирать с модели, поскольку при этом увеличивается *MSE* – несмещенная оценка остаточной дисперсии. В логарифмическом масштабе коэффициенты регрессии являются эластичностями результативного признака по каждому фактору.

Полученное уравнение регрессии значимо в целом и объясняет 99,1 % изменчивости $\ln C$, однако оценка коэффициента эластичности по прибыли оказалась больше единицы $b_1 = 1{,}18$? В последних колонках таблицы вычислены значения lnC_p и остатки модели $e = lnC - lnC_p$. График остатков модели по годам изображен на рис. 8.2.

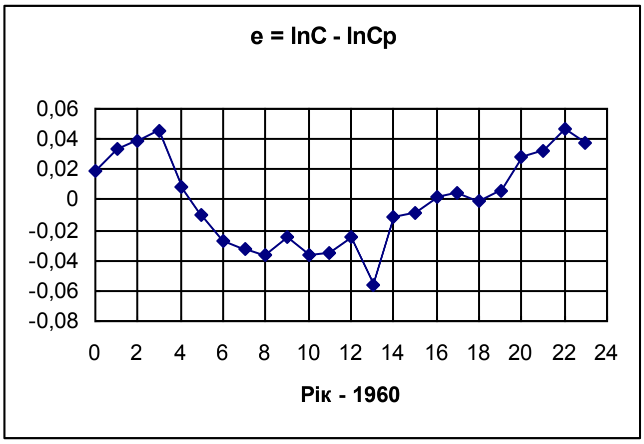


Рис. 8.2. **График остатков модели**

В данном примере на рис. 2 явно видно отсутствие случайности в поведении остатков и наблюдается эффект, похожий на наличие сезонных колебаний. Вычисленный коэффициент автокорреляции 1-го порядка $\rho = 0{,}83$ свидетельствует о наличии большой автокорреляции, это подтверждается также статистикой Дарбина-Уотсона $DW \approx 2(1 - \rho) = 0{,}33$.

Пример. В табл. 8. 3 представлены исходные данные $Y = \ln C,\ X_1 = \ln W,\ X_2 = \ln P$ и преобразованные значения $V_t = Y_t - \rho Y_{t-1},\ U_t = X_t - \rho X_{t-1}$ для $\rho = 0{,}832$.

Таблица 8.3

Данные Y, X_1, X_2 для $\rho = 0$ и V, U_1, U_2 для $\rho = 0{,}832$

Год	*Y*	*X*₀	*X*₁	*X*₂	Y_p	*e*	*V*	*U*₀	*U*₁	*U*₂	*Y_pp*	*Ee*
1960	4,159	1	6,194	4,649	4,140	0,019						
1961	4,205	1	6,222	4,655	4,171	0,033	0,743	0,168	1,066	0,785	0,726	3,479
1962	4,258	1	6,263	4,654	4,220	0,039	0,758	0,168	1,084	0,779	0,750	3,508
1963	4,304	1	6,296	4,652	4,259	0,045	0,759	0,168	1,082	0,778	0,749	3,555
1964	4,349	1	6,364	4,649	4,341	0,008	0,766	0,168	1,123	0,776	0,795	3,554
1965	4,402	1	6,424	4,644	4,412	-0,010	0,782	0,168	1,126	0,774	0,800	3,602
1966	4,446	1	6,472	4,631	4,474	-0,027	0,782	0,168	1,124	0,765	0,803	3,643
1967	4,49	1	6,512	4,627	4,523	-0,033	0,789	0,168	1,124	0,772	0,799	3,691

1968	4,538	1	6,553	4,614	4,575	-0,037	0,800	0,168	1,132	0,762	0,813	3,725
1969	4,589	1	6,583	4,605	4,613	-0,024	0,811	0,168	1,128	0,764	0,808	3,781
1970	4,625	1	6,622	4,601	4,661	-0,036	0,805	0,168	1,142	0,768	0,821	3,804
1971	4,667	1	6,658	4,605	4,702	-0,035	0,817	0,168	1,146	0,775	0,821	3,846
1972	4,723	1	6,697	4,605	4,748	-0,025	0,838	0,168	1,155	0,772	0,833	3,890
1973	4,772	1	6,763	4,596	4,828	-0,056	0,840	0,168	1,188	0,763	0,875	3,897
1974	4,822	1	6,755	4,555	4,833	-0,011	0,850	0,168	1,125	0,729	0,826	3,996
1975	4,854	1	6,775	4,536	4,864	-0,009	0,840	0,168	1,152	0,744	0,846	4,008
1976	4,905	1	6,810	4,540	4,903	0,002	0,864	0,168	1,170	0,764	0,854	4,051
1977	4,951	1	6,849	4,549	4,946	0,005	0,868	0,168	1,180	0,770	0,862	4,089
1978	5,001	1	6,896	4,551	5,001	-0,001	0,880	0,168	1,195	0,764	0,881	4,120
1979	5,042	1	6,923	4,541	5,036	0,006	0,879	0,168	1,182	0,753	0,875	4,167
1980	5,074	1	6,929	4,533	5,046	0,028	0,877	0,168	1,166	0,753	0,857	4,217
1981	5,105	1	6,956	4,545	5,073	0,032	0,881	0,168	1,188	0,772	0,870	4,235
1982	5,121	1	6,964	4,572	5,074	0,047	0,871	0,168	1,174	0,789	0,843	4,278
1983	5,143	1	6,999	4,597	5,105	0,038	0,880	0,168	1,202	0,791	0,873	4,270

По данным Y, X_1, X_2 найдены коэффициенты регрессии ($a = -1{,}783; b_1 = 1{,}190; b_2 = -0{,}312$) вычислены расчетные значения Y_p и остатки модели *e*. По этим остаткам получена начальная оценка коэффициента автокорреляции $\rho = 0{,}832$.

По преобразованным данным V, U_0, U_1, U_2 найдены новые оценки коэффициентов регрессии, вычислены новые расчетные значения Y_{pp} и новые остатки модели *ee*, по которым получена новая оценка коэффициента автокорреляции ρ^*.

В табл. 8.4 представлены результаты расчетов при разных значениях ρ . В последнем столбце значение ρ=0,999 найдено итерационно.

Таблица 8.4

Параметры модели при разных значениях ρ

ρ	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,999
a	-1,783	-1,787	-1,751	-1,663	-1,510	-1,281	-0,973	-0,588	-0,096	1,263	65,120
b_1	1,190	1,191	1,191	1,190	1,188	1,184	1,177	1,166	1,135	0,896	0,286
b_2	-0,312	-0,313	-0,321	-0,339	-0,369	-0,412	-0,470	-0,537	-0,598	-0,529	-0,410
ρ^*	0,822	0,914	0,963	0,980	0,987	0,991	0,993	0,995	0,996	0,998	0,999

Имеем, что с учетом ρ модель лишь ухудшилась. Например, для $\rho = 0{,}832$ остатки оказались одного знака и возрастают практически линейно (табл. 3 столбец *ee*) .

Данные вычисления не подтвердили эффективности итерационной процедуры Кохрана-Оркатта. Рекомендуется уточнить спецификацию модели. Есть мнение, что автокорреляция остатков является следствием неправильной спецификации модели. Например, при попытках описать нелинейную зависимость более простой линейной моделью часто появляется положительная автокорреляция остатков. На рис. 8.3 изображена линейная аппроксимация гиперболической зависимости $y = a + \frac{b}{x}$. В этом случае знаки остатков остаются одинаковыми для нескольких последовательных наблюдений, после чего опять меняются. Такое поведение остатков не есть случайным.

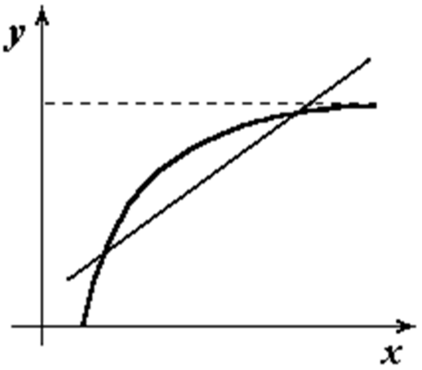


Рис. 8.3. Неправильная форма связи

График остатков этого случая очень напоминает график рис. 2. Основным источников появления нежелательных эффектов является отсутствие в модели важной объясняющей переменной и появление автокорреляции. Следовательно, необходимо не корректировать оценки параметров, а обратить внимание на возможные ошибки спецификации этой модели. При этом, выбор неправильной формы связи также является ошибкой спецификации.

Пример. Продолжим предыдущий пример. Попытка учесть автокорреляцию 1-го порядка фактически привела к модели с лаговыми переменными, т.е. с переменными по предыдущему периоду времени:

$$y_t - \rho y_{t-1} = a(1 - \rho) + b_1(x_{1t} - \rho x_{1,t-1}) + b_2(x_{2t} - \rho x_{2,t-1}) + \varepsilon_t,$$

или $$y_t = a' + b_1 x_{1t} + b_2 x_{2t} + c_0 y_{t-1} + c_1 x_{1,t-1} + c_2 x_{2,t-1} + \varepsilon_t,$$

где, согласно с традиционной методикой, коэффициенты a, b_1, b_2, c_0, c_1, c_2 связаны между собой и еще с одним параметром ρ нелинейными связями:

$$c_0 = \rho;\ c_1 = -\rho b_1;\ c_2 = -\rho b_2;\ a' = a(1 - \rho).$$

Не будем учитывать эти ограничения. В таком случае формально получим обобщенную линейную модель с лаговыми переменными, коэффициенты которой просто вычисляются обычным МНК:

$$y_{pt} = 0{,}74 + 0{,}22 x_{1t} - 0{,}20 x_{2t} + 0{,}87 y_{t-1} - 0{,}11 x_{1,t-1} + 0{,}01 x_{2,t-1}.$$

|tb| (1,5) (2,5) (1,4) (14,0) (0,9) (0,1)

В скобочках приведены значения статистик Стьюдента на каждую оценку. Последовательно выбракуем незначимые по критерию Стьюдента члены и получим такую авторегрессионную модель:

$$y_{pt} = 0{,}50 + 0{,}15x_{1t} - 0{,}16x_{2t} + 0{,}84\,y_{t-1},$$

$$\ln C_t = 0{,}50 + 0{,}15\ln W_t - 0{,}16\ln P_t + 0{,}84\ln C_{t-1}.$$

|tb| (1,3) (3,1) (2,4) (22,4)

Эта модель значимая в целом и состоит лишь со значимых по Стьюденту членов. Коэффициент детерминации $R^2 = 0{,}9996$ очень высокий, модель почти точно воспроизводит данные. Автокорреляции остатков больше нет (рис. 8.4).

Рис. 8.4. График остатков для новой модели

Новая оценка коэффициента автокорреляции $\rho = 0{,}03$ также свидетельствует об отсутствии автокорреляции. В авторегрессионной модели появился член $0{,}84 lnC_{t-1}$, что обозначает зависимость затрат текущего года lnC_t от затрат предыдущего периода. Таким образом, была найдена правильная спецификация модели.

Таким образом, была рассмотрена одна из самых возможных причин появления положительной автокорреляции остатков – это ошибки спецификации модели в виде отсутствия важных объясняющих переменных. Относительно отрицательной автокорреляции, то она не приводит к существенным ошибкам, поскольку отрицательная автокорреляция проявляется в тенденции к строгому чередованию знаков остатков модели. Но при выводе формул для вычисления дисперсий и ковариаций коэффициентов регрессии предусматривается случайное, а не строгое чередование знаков остатков. Строгое чередование знаков – это также существенная особенность и она должна быть учтена в правильной по спецификации модели.

Вопросы для самопроверки:

1. Что такое автокрреляция?
2. Какие виды автокрреляции различают?
3. В каких данных чаще всего встречается автокрреляция?
4. Какие основные причины автокрреляции?
5. Какие предпосылки МНК нарушаются при автокрреляции?
6. Какие последствия автокрреляции?
7. Перечислить основные методы обнаружения автокрреляции?
8. Описать схему использования статистики DW Дарбина-Уотсона.
9. Перечислить методы устранения автокрреляции?
10. В чем состоит суть итерационного метода Кохрана-Оркатта

Тема 9. Проблемы интерпретации параметров многофакторной модели

9.1. Интерпретация β-коэффициентов

9.2. Интерпретация параметров моделей без свободного члена

9.1. Интерпретация β-коэффициентов

Для проведения факторного анализа в экономике чаще всего используют линейную множественную регрессию, цель которой – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель [31, с. 91]. Уравнение линейной множественной регрессии имеет вид:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_mx_m, \tag{1}$$

где $\hat{y}$ – результативный признак (зависимая переменная);

$x_1, x_2, \ldots, x_m$ – факторы (независимые переменные);

$b_0, b_1, b_2, \ldots, b_m$ – параметры регрессии.

Остановимся на интерпретации параметров регрессии. Специалисты в области корреляционно-регрессионного анализа считают, что постоянная b_0 выполняет в уравнении регрессии функцию выравнивания и определяет точку пересечения гиперповерхности регрессии с осью ординат [30, с. 76]. Значения $b_1, b_2, \ldots, b_m$ –

оценки коэффициентов регрессии. Параметр b_i указывает среднюю величину изменения y при изменении x_i на одну единицу при условии, что другие переменные остаются без изменения. В реальных моделях отрицательные значения коэффициентов b_i вызывают недоумение в их интерпретации, поскольку часто имеется противоречие действующим в экономике закономерностям взаимосвязи между показателями. Объяснением такого факта есть малое число наблюдений. В любом случае рекомендуется дополнить уравнение чистой регрессии (1) уравнением регрессии в стандартизированных величинах:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m, \qquad (2)$$

где $\hat{Y}, X_1, X_2, X_m$ – стандартизированные переменные;

$\beta_1, \beta_2, \beta_m$ – стандартизированные коэффициенты регрессии.

Тут появляется возможность сравнения действия факторов на результативный признак. Сравнение происходит по оценке интенсивности влияния факторов на результативный признак, т.е. по величине β_i. Стандартизированные коэффициенты β_i множественной регрессии характеризуют скорость изменения среднего значения результативного признака по каждому фактору при постоянных значениях остальных переменных, включенных в анализ. Другими словами, на какую часть стандартного отклонения изменилось бы среднее значение Y, если бы значение соответствующего X_i увеличилось на стандартное отклонение, а остальные переменные остались без изменения. При этом β_i принимают значения в интервале [-1, 1]. Однако в решении реальных экономических задачах они иногда выходят за пределы этого интервала. Традиционно считается, что этого не может быть, и когда на практике встречается указанный эффект, его объясняют возможными, пока не обнаруженными ошибками в компьютерной программе.

Впервые мнение о невозможности появления бета-коэффициентов, больших единицы по модулю, было высказано в книге [36]. Основание для такого вывода было приведено всего одно – в одномерном случае (одна объясняющая переменная) стандартизованный коэффициент регрессии совпадает с парным коэффициентом корреляции, который, естественно, не может быть больше единицы по абсолютной величине. Приведенное обоснование не является достаточным для того, чтобы делать какие-либо заключения о поведении β коэффициентов в многомерном случае. Однако такое мнение существует до сих пор.

Рассмотрим эту проблему для двумерного случая, т.е. с двумя факторами. В работе [5] показано, что при некоторых сочетаниях коэффициентов корреляции β могут быть больше единицы по абсолютной величине. Например, для таких коэффициентов корреляции $r_{x_1 x_2} = 0{,}7;$ $r_{x_1 y} = -0{,}35;$ $r_{x_2 y} = 0{,}35$ в результате решения системы нормальных уравнений

$$\begin{cases} r_{x_1 y} = \beta_1 + \beta_2 r_{x_1 x_2} \\ r_{x_2 y} = \beta_1 r_{x_1 x_2} + \beta_2 \end{cases} \qquad \begin{cases} -0{,}35 = \beta_1 + 0{,}7\beta_2 \\ 0{,}35 = 0{,}7\beta_1 + \beta_2 \end{cases}$$

получается $\beta_1 = -1{,}167$; $\beta_2 = 1{,}167$ (т.е. оба стандартизованных коэффициента регрессии оказываются большими единицы по модулю). Дополнительно вычисляем коэффициент детерминации и коэффициент множественной корреляции:

$$R^2 = \beta_1 r_{x_1 y} + \beta_2 r_{x_2 y} = (-1{,}167)\cdot(-0{,}35) + 1{,}167 \cdot 0{,}35 = 0{,}817;$$

$$R = \sqrt{0{,}817} = 0{,}904\,.$$

Два фактора x_1 и x_2 объясняют около 82% изменчивости результативной переменной y.

Можно вычислить еще коэффициенты частной корреляции, когда один из факторов фиксируется на своем среднем значении:

$$r_{yx_1 . x_2} = \beta_1 \cdot \sqrt{\frac{1 - r^2_{x_1 x_2}}{1 - r^2_{yx_2}}} = -1{,}167 \cdot \sqrt{\frac{1 - 0{,}7^2}{1 - 0{,}35^2}} = -0{,}889;$$

$$r_{yx_2 . x_1} = \beta_2 \cdot \sqrt{\frac{1 - r^2_{x_1 x_2}}{1 - r^2_{yx_1}}} = 1{,}167 \cdot \sqrt{\frac{1 - 0{,}7^2}{1 - 0{,}35^2}} = 0{,}889.$$

Объяснение значений β, вышедших за пределы интервала [-1, 1] следующее. После стандартизации значения всех переменных образуют систему многомерных векторов одинаковой "длины" – их скалярные квадраты одинаковы и равны n (числу наблюдений). Можно нормировать все эти векторы так, чтобы их длины равнялись единице:

$$Y = \frac{y - \overline{y}}{s_y \sqrt{n}}; \quad X_1 = \frac{x_1 - \overline{x_1}}{s_{x_1} \sqrt{n}}; \quad X_2 = \frac{x_2 - \overline{x_2}}{s_{x_2} \sqrt{n}},$$

где s_y, s_{x_1}, s_{x_2} – средние квадратичные отклонения:

$$s_y = \sqrt{\frac{1}{n}\sum (y - \overline{y})^2}; \quad s_{x_1} = \sqrt{\frac{1}{n}\sum (x_1 - \overline{x_1})^2}; \quad s_{x_2} = \sqrt{\frac{1}{n}\sum (x_2 - \overline{x_2})^2}.$$

Теперь все скалярные квадраты равны единице $(Y, Y) = 1$; $(X_1, X_1) = 1$; $(X_2, X_2) = 1$, а скалярные произведения разных векторов равны коэффициентам корреляции

$$(X_1, X_2) = r_{x_1 x_2}; \quad (X_1, Y) = r_{yx_1}; \quad (X_2, Y) = r_{yx_2}.$$

Уравнение регрессии в этих переменных

$$Y = Yp + E = \beta_1 X_1 + \beta_2 X_2 + E$$

можно трактовать как разложение единичного вектора Y по системе координат (X_1, X_2, E), где вектор E – вектор ошибок (невязок), который ортогонален плоскости единичных векторов (X_1, X_2): $E \perp X_1$, $E \perp X_2$, $E \perp Yp$, так как согласно методу наименьших квадратов (МНК) коэффициенты регрессии определяются из системы "нормальных" уравнений, в которой записаны условия "нормальности", ортогональности вектора ошибок к каждому члену модели (т.е. к каждому вектору факторов). Вектор расчетных значений $Yp = \beta_1 X_1 + \beta_2 X_2$ есть ортогональная проекция единичного вектора Y на плоскость векторов (X_1, X_2). Скалярный квадрат этого вектора равен коэффициенту детерминации $(Yp, Yp) = R^2$, а его длина – коэффициенту множественной корреляции $|Yp| = R \le 1$. При такой интерпретации стандартизованные коэффициенты регрессии β_1, β_2 будут косоугольными проекциями в разложении вектора Yp (его длина не больше единицы) по системе единичных, но не ортогональных векторов X_1, X_2. Угол между базисными векторами X_1, X_2 равен $arccos(r_{x_1 x_2})$. Например, для $r_{x_1 x_2} = 0{,}7$ этот угол равен 0,795 радиан или 45,57°. Возникает вопрос, могут ли проекции единичного вектора в разложении по единичному базису быть больше единицы по абсолютной величине? Нет, если базис ортогональный и возможно, если базис не ортогональный.

На рис. 9.1 изображено разложение вектора Yp по базису X_1, X_2 для $r_{x_1 x_2} = 0{,}7$; $r_{yx_1} = -0{,}35$; $r_{yx_2} = 0{,}35$. Для принятых значений коэффициентов корреляции угол между базисными векторами равен 45,57°; длина вектора расчетных значений равна $R = 0{,}904$; косоугольные проекции этого вектора по единичному базису X_1, X_2 оказались по модулю большими единицы: $\beta_1 = -1{,}167$; $\beta_2 = 1{,}167$.

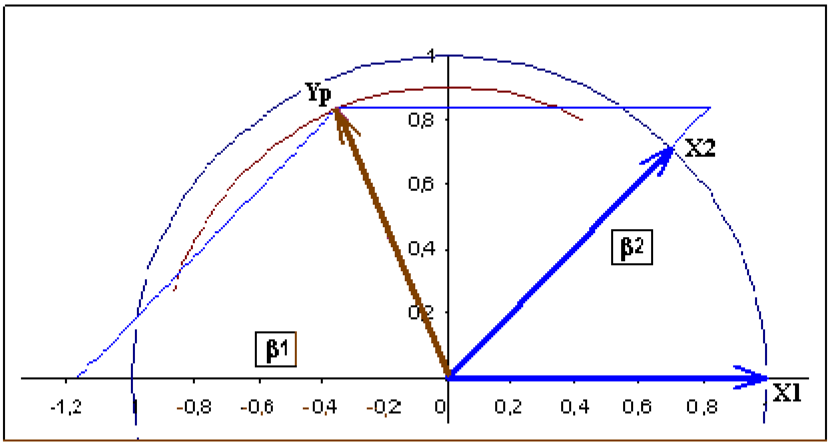


Рис. 9.1. Косоугольные координаты вектора *Yp* по базису X_1, X_2

Математически здесь все правильно. Рассмотрим реальный пример с экономического анализа зависимости эффекта финансового рычага (*У*) от коэффициента износа основных средств (X_1), коэффициента обновления основных средств (X_2), коэффициента выбытия основных средств (X_3), коэффициента оборачиваемости основных средств (X_4), коэффициента общей (текущей) ликвидности (X_5), коэффициента срочной ликвидности (X_6), коэффициента абсолютной ликвидности (X_7), коэффициента оборачиваемости активов (X_8), коэффициента оборачиваемости дебиторской задолженности (X_9), периода оборачиваемости дебиторской задолженности (X_{10}), коэффициента оборачиваемости запасов (X_{11}), длительности оборачиваемости запасов (X_{12}), коэффициента оборачиваемости совокупного капитала (X_{13}). В табл. 9.1 приведены значения этих показателей.

Таблица 9.1

Значения финансовых показателей

x1	x2	x3	x4	x5	x6	x7	X8	x9	x10	x11	x12	x13	У
0,61	0,12	0,14	3,02	9,65	5,59	2,32	1,07	4,23	86,37	3,73	97,91	5,85	0,21
0,61	0,14	0,10	3,23	10,65	6,62	3,14	0,91	3,67	99,59	3,35	109,04	5,40	0,10
0,59	0,17	0,08	2,98	12,79	8,31	3,59	0,78	2,99	122,08	3,66	99,76	4,69	0,08
0,66	0,01	0,10	1,49	2,39	1,56	0,21	1,10	7,32	49,85	9,22	39,58	2,01	0,06
0,72	0,01	0,00	1,47	13,89	8,79	1,62	0,95	4,38	83,31	4,49	81,21	2,20	0,03
0,52	0,17	0,10	5,69	0,92	0,43	0,02	2,16	21,72	16,81	9,56	38,17	5,21	0,24
0,91	0,01	0,00	7,58	1,76	0,91	0,01	2,62	11,51	31,72	9,26	39,42	9,12	0,11
0,89	0,04	0,00	8,97	1,27	0,51	0,00	3,38	17,96	20,32	10,74	33,99	9,06	0,12
0,37	0,13	0,02	2,08	2,61	1,98	0,06	1,39	8,69	42,00	-42,19	-8,65	3,10	0,25
0,41	0,09	0,01	4,03	2,46	1,57	0,12	1,94	13,31	27,43	-67,74	-5,39	5,64	0,27
0,43	0,13	0,00	4,98	3,26	2,68	0,29	2,29	12,63	28,91	-58,31	-6,26	7,32	0,48
0,44	0,15	0,07	8,18	4,44	3,75	0,89	2,87	15,15	24,09	-28,54	-12,79	11,29	0,32

0,71	0,02	0,00	1,11	5,30	2,34	1,42	1,08	21,42	17,04	5,85	62,38	1,95	0,13
0,76	0,02	0,23	1,31	6,23	4,80	0,16	0,76	5,26	69,40	-6,17	-59,14	2,43	0,11
0,74	0,02	0,16	1,95	4,65	3,19	0,25	0,92	4,98	73,35	-21,85	-16,71	3,09	0,17
0,73	0,05	0,08	2,76	5,74	3,98	0,05	1,13	4,47	81,68	9,52	38,33	3,23	0,07
0,63	0,02	0,01	0,73	2,53	1,44	0,36	0,58	4,19	87,13	2,29	159,38	1,17	0,03

Построив зависимости результативного признака $У$ от факторов, получили модели у которых β коэффициенты вышли за пределы интервала [-1, 1]:

$$У = \beta_4 X_4 + \beta_8 X_8 \ ;$$

$$У = \beta_4 X_4 + \beta_{13} X_{13};$$

$$У = \beta_5 X_5 + \beta_6 X_6 \ ;$$

$$У = \beta_4 X_4 + \beta_8 X_8 + \beta_{13} X_{13} .$$

В табл. 9.2 приведены коэффициенты парной корреляции для переменных $У$, X_4, X_8, X_{13}, X_5, X_6.

Таблица 9.2

Корреляционная матрица

	У	***X*₄**	***X*₈**	***X*₁₃**	***X*₅**	***X*₆**
У	1	0,409861	0,510128	0,517883	-0,34474	-0,27375
X_4	0,409861	1	0,94595	0,945351	-0,37805	-0,351
X_8	0,510128	0,94595	1	0,861387	-0,548	-0,5232
X_{13}	0,517883	0,945351	0,861387	1	-0,21969	-0,17953
X_5	-0,34474	-0,37805	-0,548	-0,21969	1	0,982608
X_6	-0,27375	-0,351	-0,5232	-0,17953	0,982608	1

Для первой модели имеем $r_{x_4 x_8} = 0{,}946;$ $r_{yx_4} = 0{,}410;$ $r_{yx_8} = 0{,}510$. В результате расчетов получаем $\beta_4 = -0{,}691$; $\beta_2 = 1{,}164$; $R = 0{,}557$. Один из β-коэффициентов оказался большим единицы. На рис. 2 приведена геометрическая интерпретация этой ситуации.

Для второй модели имеем $r_{x_4 x_{13}} = 0{,}945;$ $r_{yx_4} = 0{,}410;$ $r_{yx_{13}} = 0{,}518$. Эти значения мало отличаются от аналогичных значений предыдущей модели. В результате расчетов получаем $\beta_4 = -0{,}750$; $\beta_{13} = 1{,}227$; $R = 0{,}573$. Здесь также один из β-коэффициентов оказался большим единицы. Геометрическая интерпретация очень похожа на рис. 9.2.

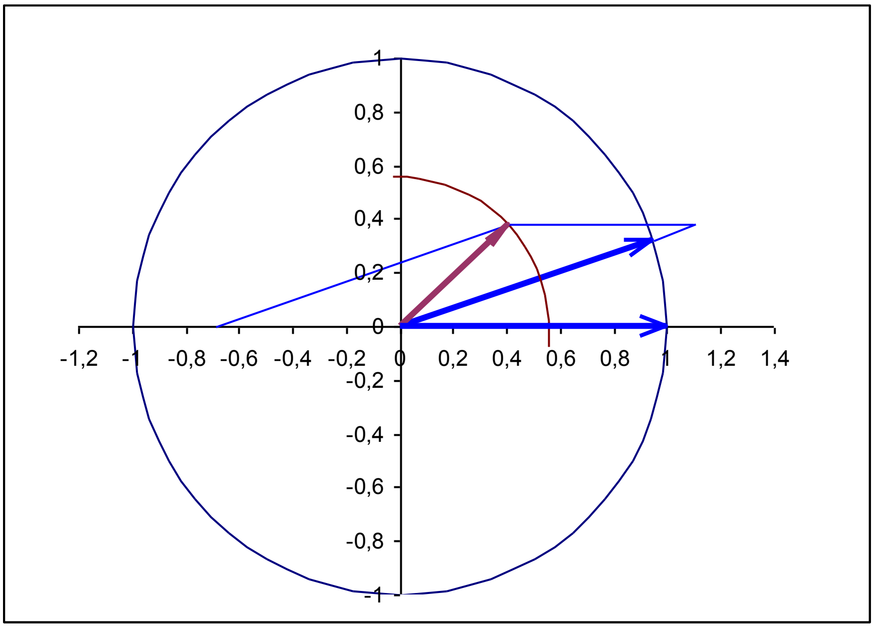


Рис. 9.2. Геометрическая интерпретация для первой модели

Для третьей модели $r_{x_5x_6} = 0{,}983$; $r_{yx_5} = -0{,}345$; $r_{yx_6} = -0{,}274$, откуда получаем $\beta_5 = -2{,}197$; $\beta_6 = 1{,}885$; $R = 0{,}491$. Здесь оба β-коэффициенты оказались значительно больше единицы по абсолютной величине. Геометрическая интерпретация приведена на рис. 9.3.

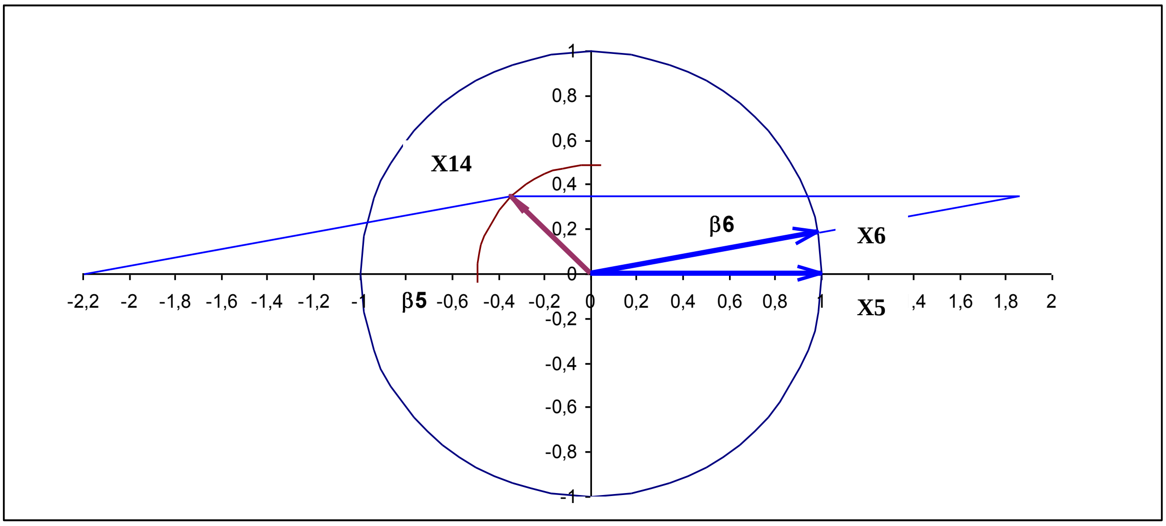


Рис. 9.3. Геометрическая интерпретация для третьей модели

Теснота корреляционных связей для всех вышеприведенных двумерных моделей была невысокой с коэффициентами множественной корреляции $R = 0{,}49 \div 0{,}57$ (модели объясняли всего от $R^2 \approx 24\%$ до $R^2 \approx 32\%$ полной изменчивости).

Кроме анализа двумерных моделей была составлена многомерная модель, в которую были включены все вышеуказанные факторы. Алгоритм пошаговой регрессии выбраковал как незначимые переменные x_5, x_6 и в результате была получена трехмерная модель $У = \beta_4 X_4 + \beta_8 X_8 + \beta_{13} X_{13}$ с коэффициентом множественной корреляции $R = 0{,}779$ (иными словами, трехмерная модель объяснила около $R^2 \approx 61\%$ полной изменчивости данных). Модель значима в целом и состоит только из значимых членов. Все β коэффициенты и для этого случая оказались большими единицы по абсолютной величине.

В табл. 9.3 приведен фрагмент корреляционной матрицы из табл. 9.2 для переменных $У$, X_4, X_8, X_{13}.

Таблица 9.3

Корреляционная матрица для $У$, X_4, X_8, X_{13}

	$У = X_{14}$	X_4	X_8	X_{13}
X_4	0,409861	1	0,94595	0,945351
X_8	0,510128	0,94595	1	0,861387
X_{13}	0,517883	0,945351	0,861387	1

Стандартизованные коэффициенты регрессии определяются из системы нормальных уравнений:

$$0{,}410 = \beta_4 + 0{,}946\cdot\beta_8 + 0{,}945\cdot\beta_{13}$$
$$0{,}510 = 0{,}946\cdot\beta_4 + \beta_8 + 0{,}861\cdot\beta_{13}$$
$$0{,}518 = 0{,}945\cdot\beta_4 + 0{,}861\cdot\beta_8 + \beta_{13}$$

Решением данной системы являются числа $\beta_4 = -2{,}870$; $\beta_8 = 1{,}713$; $\beta_{13} = 1{,}756$, т.е. все β-коэффициенты оказались существенно большими единицы по абсолютной величине. Коэффициент детерминации и коэффициент множественной корреляции:

$$R^2 = -2{,}870\cdot 0{,}410 + 1{,}713\cdot 0{,}510 + 1{,}756\cdot 0{,}518 = 0{,}606;$$
$$R = 0{,}779.$$

Значимость модели в целом:

$F = \dfrac{R^2}{1-R^2}\cdot\dfrac{m}{n-m-1} = \dfrac{0{,}606}{1-0{,}606}\cdot\dfrac{3}{17-3-1} = 6{,}685$, значение больше табличного $F_{0,01}(3;13) = 5{,}74$, откуда следует вывод о значимости модели в целом.

Уравнение зависимости эффект финансового рычага ($У$) от коэффициента оборачиваемости основных средств (X_4), коэффициента оборачиваемости активов (X_8), коэффициента оборачиваемости совокупного капитала (X_{13}) (у каждого коэффициента регрессии в скобках приведены значения статистик Стьюдента):

$$У = -2{,}871 \cdot X_4 + 1{,}713 \cdot X_8 + 1{,}756 \cdot X_{13}\,.$$
$$(-3{,}3) \qquad (3{,}0) \qquad (3{,}1)$$

Таким образом, модель значима в целом (по критерию Фишера) и состоит только из значимых (по критерию Стьюдента) членов, однако остаются β коэффициенты большие единицы по модулю.

В современных статистических пакетах обычно сразу получают уравнение чистой регрессии:

$$y = -0{,}0733 - 0{,}133 \cdot x_4 + 0{,}242 \cdot x_8 + 0{,}0717 \cdot x_{13}\,.$$
$$(-1{,}3) \quad (-3{,}3) \qquad (3{,}0) \qquad (3{,}1)$$

Согласно модели при увеличении коэффициента оборачиваемости основных средств (X_4) значение эффекта финансового рычага уменьшается на 0,133 единицы. В этом примере явно видна ошибочность традиционной трактовки коэффициентов регрессии. Действительно, здесь все три фактора связаны очень тесными корреляционными связями $r_{x_4 x_8} = 0{,}946;$ $r_{x_4 x_{13}} = 0{,}945;$ $r_{x_8 x_{13}} = 0{,}861$ и поэтому они не могут варьировать независимо друг от друга. Фактически это разные формы одного и той же причины с положительной корреляцией с результативным признаком $r_{yx_4} = 0{,}410;$ $r_{yx_8} = 0{,}510$ $r_{yx_{13}} = 0{,}518$. Таким образом, традиционная интерпретация справедлива лишь при независимости объясняющих переменных, когда каждый аргумент можно изменяться независимо от других. В регрессионном анализе, как правило, факторы не являются независимыми, между ними почти всегда есть корреляционные связи, именно поэтому для определения коэффициентов регрессии составляется и решается система нормальных уравнений.

Уравнение регрессии представляет собой компактное описание исходных данных. Методом наименьших квадратов получено выражение с наименьшими ошибками, которое предназначено для наилучшего воспроизведения данных. Строя же модели в экономическом анализе, происходит использование данного выражения не совсем по назначению и при этом делаются выводы при варьировании факторов в других диапазонах по сравнению с исходными данными. Да, тут появляется аргумент – отсутствие альтернативных моделей.

В задаче определении зависимости эффекта финансового рычага ($У$) от факторов с тесными мультиколлинеарными связями проблема разрешается переходом в систему главных компонент. В табл. 9.4 вычислены собственные числа и собственные векторы для корреляционной матрицы 3-го порядка относительно X_4, X_8, X_{13} :

Таблица 9.4

Собственные числа λ и собственные векторы U_1, U_2, U_3

№	U_1	U_2	U_3
1	0,589	–0,002	0,808
2	0,571	–0,706	–0,418
3	0,571	0,708	–0,414
λ	2,836	0,139	0,026
%	94,5%	4,6%	0,9%
Σ	94,5%	99,1%	100%

Компоненты составляются как линейные комбинации исходных стандартизованных переменных с коэффициентами, записанными в собственных векторах. Если округлить цифры в табл. 4, получается, что компоненты в этом примере будут близки к таким выражениям:

$$F_1 = X_4 + X_8 + X_{13}$$

$$F_2 = - X_8 + X_{13}$$

$$F_3 = 2 \cdot X_4 - X_8 - X_{13}$$

Компоненты F_1, F_2, F_3 взаимно ортогональны (независимы), центрированы, но не нормированы. Нормированные (стандартизованные) компоненты обозначим Φ_1, Φ_2, Φ_3. При этом первая (главная) компонента Φ_1 будет близка к среднему арифметическому исходных стандартизованных переменных.

В модель следует включать не все, а только главные компоненты с дисперсиями (собственными числами), большими единицы. Кроме того, в сумме главные компоненты должны объяснять около 80% общей изменчивости. В нашей задаче этим требованиям удовлетворяет одна первая (главная) компонента, которая объясняет 94,5% полной изменчивости. Остальные две компоненты фактически не варьируют и представляют собой малые ошибки.

При учете в модели только главных компонент отбрасываются малые ошибки, помимо всех дальнейших выбраковок, которые могут быть сделаны на основании статистических критериев.

Имеем следующее уравнение:

$$У = 0{,}492 \cdot \Phi_1 = 0{,}172 \cdot X_1 + 0{,}167 \cdot X_2 + 0{,}167 \cdot X_3,$$

которое объясняет всего $R^2 = 0{,}492^2 \approx 24{,}2\%$ изменчивости $У$. По сравнению с одномерными моделями

$$У = 0{,}410 \cdot X_1,$$
$$У = 0{,}510 \cdot X_2,$$
$$У = 0{,}518 \cdot X_3,$$

в которых весь совокупный эффект приписывается одному фактору, в трехфакторной модели (полученной с помощью метода главных компонент) совокупный эффект разделен на три примерно одинаковые части.

Существует нежелательный эффект включения в модель малых ошибок, после чего резко меняются значения коэффициентов регрессии ("прохлопывание"), что иногда можно предотвратить, если при последовательном подключении очередной переменной анализировать остаточные дисперсии каждой переменной, еще не включенной в модель [8].

Однако, алгоритм последовательного подключения–исключения не доведен до завершения, поскольку он никак не учитывает возможное исчерпывание изменчивости объясняющих переменных. Существующее стандартное программное обеспечение не гарантирует предотвращения бессмысленных результатов.

Предположение, что факт $|\beta| > 1$ связан с мультиколлинеарностью дополняется предположением о неправильном определение направления причинно-следственных связей, когда в качестве результативной выбрана не та переменная, другими словами принята неверная спецификация модели.

Если допустить, что результативным признаком может быть любой из факторов (табл. 1), то можно найти 66 двухфакторных моделей, для которых β коэффициенты будут больше единицы по модулю. В табл. 9.5 найденные комбинации $(У, x_1, x_2)$ расположены по столбцам в порядке возрастания коэффициента множественной корреляции R (от 0,29 для первой комбинации до 0,99 для последней 66-й комбинации). При этом уменьшался коэффициент корреляции между факторами $r_{x_1x_2}$ (от 0,98 до 0,25). Выделены комбинации, для которых хотя бы один из β коэффициентов оказался больше 2 (по модулю). Для комбинации (12, 5, 6) оба β коэффициенты оказались большими 3 (по модулю).

Таблица 9.5

Комбинации ($У$, x_1, x_2) с $|\beta| > 1$

(2, 5, 6)	(11, 9, 10)	(12, 9, 10)	(9, 8, 13)	**(10, 4, 8)**	(4, 8, 9)
(13, 5, 6)	**(11, 5, 6)**	(6, 4, 8)	(5, 9, 10)	(9, 10, 12)	(13, 4, 6)
(1, 5, 6)	(3, 4, 8)	(5, 4, 8)	(2, 1, 11)	(13, 8, 10)	(4, 5, 8)
(1, 4, 13)	(7, 4, 13)	(7, 4, 8)	(7, 9, 10)	(13, 7, 8)	(4, 6, 8)
(3, 5, 6)	(14, 4, 8)	**(12, 5, 6)**	(10, 8, 13)	(9, 7, 10)	(4, 7, 8)
(2, 4, 8)	(5, 4, 13)	(5, 8, 13)	(7, 5, 6)	(4, 13, 15)	(4, 8, 10)
(11, 4, 13)	(14, 4, 13)	(6, 8, 13)	(13, 8, 9)	(13, 4, 10)	(5, 6, 9)
(2, 4, 13)	(6, 4, 13)	(7, 8, 13)	(9, 6, 10)	(13, 4, 8)	(6, 5, 15)
(11, 4, 8)	(10, 4, 13)	(9, 4, 8)	(9, 5, 10)	(8, 4, 13)	(6, 5, 7)
(14, 5, 6)	(9, 4, 13)	(6, 9, 10)	(13, 5, 8)	(13, 4, 9)	(6, 5, 11)
(12, 4, 8)	(9, 5, 6)	(11, 1, 2)	(13, 6, 8)	(13, 4, 5)	(6, 5, 12)

Среди этих комбинаций интерес представляют X_6, X_5, X_{11}, где в качестве результативной выбрана переменная $У = X_6$. Мультиколлинеарности здесь нет, так как $r_{x_5 x_{11}} = 0{,}252$. В табл. 5 имеется 15 комбинаций с малыми коэффициентами корреляции $|r_{x_i x_j}| < 0{,}6$, и в всех этих моделях присутствует изучаемый эффект. Таким образом, установлено, что появление β коэффициентов больших по модулю единицы не связано с наличием тесных мультиколлинеарных связей.

Во всех найденных комбинациях неприятный эффект исчезает, если заменить результативную переменную на другую из этого же списка. Так, на рис. 3 изображена геометрическая интерпретация для X_5, X_6, X_{14}, где в качестве результативной выбрана последняя переменная $У = X_{14}$. Оба стандартизованных коэффициента регрессии оказались существенно большими единицы (по модулю): $\beta_5 = -2{,}197$; $\beta_6 = 1{,}885$. Теснота связи для этой модели – невысокая $R = 0{,}491$. Если же принять в качестве результативной переменную X_5, получается доброкачественная модель $\beta_6 = 0{,}960$; $\beta_{15} = -0{,}082$ с высокой теснотой связи $R = 0{,}986$. На рис. 9.4 изображена геометрическая интерпретация результатов этих расчетов.

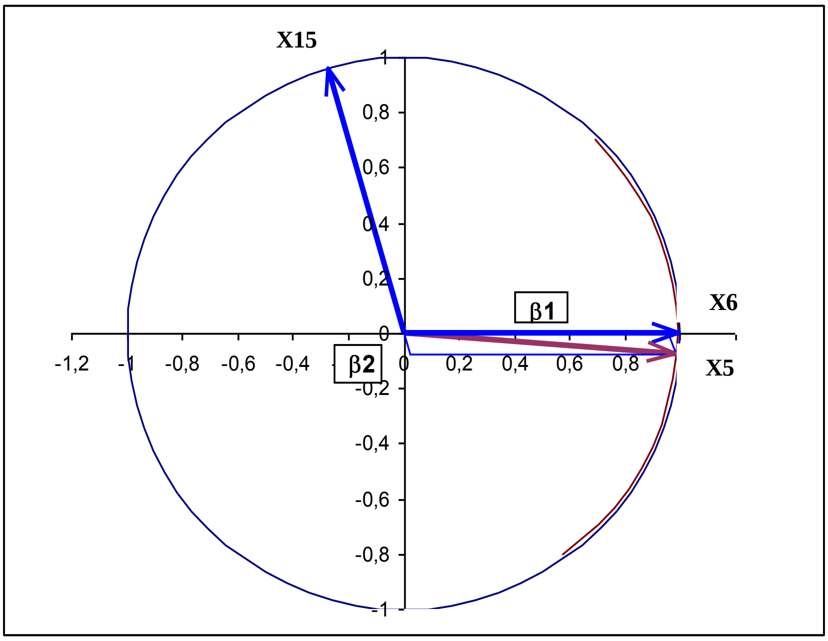


Рис. 9.4. Модель $X_5 = \beta_6 X_6 + \beta_{15} X_{15}$

Метод наименьших квадратов является основным инструментом для описания корреляционных связей. Альтернативы этому методу пока не предложено. Однако, в регрессионном анализе за исследователем остается выбор результативного признака. При этом все случайные (и неслучайные) ошибки будут отнесены только к выбранной результативной переменной, а остальные (объясняющие) переменные будут считаться измеренными точно. Ошибочное представление структуры модели при неверном принятии некоторой переменной как результативной может иметь самые различные последствия. Появление значений β коэффициентов, больших единицы по модулю, – одно из таких последствий, и наличие этого эффекта надо расценивать как явный признак неверного определения структуры (спецификации) модели.

Таким образом, вычисляя регрессионные модели для экономического анализа следует знать, что может получится уравнение с β коэффициентами по модулю большими единицы. Причем статистические критерии значимости отдельных членов модели подтвердят ее качество; они не предотвратят включение в модель посторонних помех, в результате чего может быть получено уравнение регрессии, лишенное какого-либо смысла. При наличии мультиколлинеарности можно рекомендовать обязательное применение метода главных компонент для предотвращения нежелательных искажений результатов регрессионного анализа. Однако при получении модели с β коэффициентами по модулю большими единицы рекомендуется тщательно изучить направление причинно-следственных связей изучаемых явлений и уточнить спецификацию модели.

9.2. Интерпретация параметров моделей без свободного члена

Рассмотрим одну из проблем множественной регрессионной модели, связанной с содержанием свободного члена. Свободный член линейной многофакторной модели $y = b_0 + b_1 x_1 + b_2 x_2$ играет ту же самую роль, что и в однофакторном случае – он нужен в модели для того, чтобы при средних хначениях объясняющих переменных ($\bar{x}_1, \bar{x}_2$) было получено среднее значение результативного признака ($\bar{y}$).

Примеров регрессионных моделей, у которых свободный член равен нулю много в экономике: гипотеза непрерывной (постоянной) прибыли Милтона Фридмана, согласно которой постоянное потребление пропорционально постоянной прибыли; в теории анализа затрат утверждается, что переменные затраты производства пропорциональны выпуску продукции; в монетарной теории уровень изменения цен то есть уровень инфляции пропорциональный уровню изменения предложения денег [1]. З позиции математики модели без свободного члена появляются в случаи, когда известно, что линия регрессии обязательно должна проходить через фиксированный узел, чаще всего через начало координат. Например, представленная на рис. 9.5 квадратическая зависимость всегда проходит через начало координат при любых значениях параметров: $y = b_1 x + b_2 x^2$. Модели без свободного члена также появляются при использовании взвешенного метода наименьших квадратов, когда для преодоления последствий гетероскедастичности все уравнения первоначально умножаются на весовую функцию. Например, нелинейную модель $y = \dfrac{1}{b_0 + b_1 x}$ можно привести к линейному виду (относительно параметров) функциональным преобразованием зависимой переменной ${}^1\!/_y = b_0 + b_1 x$. Однако известно, что при функциональных преобразованиях зависимой переменной почти всегда появляется существенная гетероскедастичной остатков модели, способная исказить значения оценок параметров модели. Весовая функция предназначена придать больший вес

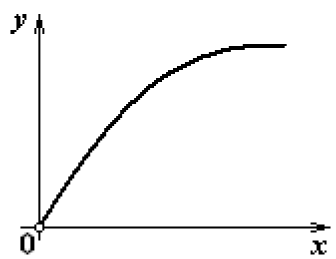


Рис. 9.5. Квадратическая модель

надежным наблюдениям и меньший нестабильным данным с большой изменчивостью. В данном примере весовая функция равна $g = y^2$. Умножим линейное уравнение (в координатах x, $^1/_y$) на эту весовую функцию и получим модель без свободного члена $y = b_0 y^2 + b_1 x y^2$, которая линейная относительно параметров и не имеет гетероскедастичности. МНК-оценки параметров такой модели – доброкачественные и не имеют систематических ошибок (смещений).

Однако отсутствие в модели свободного члена приводит к тому, что сумма остатков не будет равняться нулю. Все статистические характеристики качества модели – коэффициент детерминации, статистики Фишера и Стьюдента, вычисленные по стандартным формулам будут теперь неправильными. Модели с отсутствующим или нулевым пересечением можно использовать в некоторых случаях, однако здесь необходимо помнить несколько специфических моментов. Во-первых, $\sum e_i$, которая всегда равна нулю в модели с имеющейся величиной пересечения (общепринятая модель), не всегда должна равняться нулю для регрессии, проходящей через начало координат. Во-вторых, R^2 - коэффициент детерминации, который всегда положительный в общепринятой модели, в некторых случаях может преобразоваться в отрицательный для регрессии, проходящей через начало координат. Этот случай мы получим потому, что R^2 неоднозначно предусматривает, что пересечение включается в модель. Поэтому R^2, полученный по правилам, может не отвечать регрессионной модели, которая проходит через начало координат.

Именно из-за этих специфических особенностей модели регрессии, проходящей через начало координат, пользоваться следует осторожно. Если нет достаточной первоначальной уверенности в ее использовании, лучше использовать общепринятую модель с имеющимся пересечением. Это имеет дважды преимущества. Во-первых, если мы включаем в модель величину пересечения, но она статистически не значима для всех практических целей, имеем регрессию, проходящую через начало координат. Во-вторых, и это главное, если фактически пересечение существует, но мы настаиваем на том, чтобы использовать регрессию, проходящую через начало координат, то мы допускаем ошибку спецификации, тем самым нарушая предпосылку 5 классической модели линейной регрессии [15, с. 133 – 134].

Рассмотрим детально модель без свободного члена. Для оценки параметров моделей без свободного члена $y = b_1 x_1 + b_2 x_2 + e$ предлагается добавить в метод наименьших квадратов обязательное условие $\sum e = 0$. Имеем задачу на условный экстремум: необходимо найти минимум суммы квадратов остатков модели при условии равенства нулю их суммы:

$$\begin{cases} \sum e^2 \to \min, \\ \sum e = 0. \end{cases}$$

Составим функцию Лагранжа, где множитель Лагранжа обозначено 2λ:

$$F = \sum e^2 - 2\lambda \sum e$$

и приравняем нулю ее частные производные по всем переменным λ, b_1, b_2:

$$\begin{cases} \frac{\partial F}{\partial \lambda} = 2\sum e = 0 \\ \frac{\partial F}{\partial b_1} = 2\sum e \frac{\partial e}{\partial b_1} - 2\lambda \sum \frac{\partial e}{\partial b_1} = 0 \\ \frac{\partial F}{\partial b_2} = 2\sum e \frac{\partial e}{\partial b_2} - 2\lambda \sum \frac{\partial e}{\partial b_2} = 0 \end{cases}.$$

Учтем, что

$$e = y - b_1 x_1 - b_2 x_2,$$

$$\frac{\partial e}{\partial b_1} = -x_1; \quad \frac{\partial e}{\partial b_2} = -x_2.$$

Откуда получим такую систему связей по остаткам модели:

$$\begin{cases} \sum e = 0, \\ \sum e x_1 = \lambda \sum x_1, \\ \sum e x_2 = \lambda \sum x_2. \end{cases}$$

Умножим уравнение $y = b_1 x_1 + b_2 x_2 + e$ на каждую переменную $(1, x_1, x_2, y, e)$ и просуммируем эти выражения по всем наблюдениям:

$$\sum y = b_1 \sum x_1 + b_2 \sum x_2 + \sum e,$$
$$\sum y x_1 = b_1 \sum x_1 x_1 + b_2 \sum x_1 x_2 + \sum e x_1,$$
$$\sum y x_2 = b_1 \sum x_1 x_2 + b_2 \sum x_2 x_2 + \sum e x_2,$$
$$\sum y y = b_1 \sum y x_1 + b_2 \sum y x_2 + \sum y e,$$
$$\sum y e = b_1 \sum e x_1 + b_2 \sum e x_2 + \sum e e.$$

Учтем связи, положенные на остатки и с первых трех суммарных выражений получим такую систему уравнений для оценки параметров модели без свободного члена:

$$\begin{cases} \sum y = b_1 \sum x_1 + b_2 \sum x_2, \\ \sum y x_1 = \lambda \sum x_1 + b_1 \sum x_1 x_1 + b_2 \sum x_1 x_2, \\ \sum y x_2 = \lambda \sum x_1 + b_1 \sum x_1 x_2 + b_2 \sum x_2 x_2. \end{cases}$$

Сравним полученную систему уравнений с системой «нормальных уравнений» для оценки параметров модели со свободным членом $y = b_0 + b_1 x_1 + b_2 x_2$:

$$\begin{cases} \sum y \;\; = nb_0 \quad\;\; b_1\sum x_1 \;\; + b_2\sum x_2 \;\;, \\ \sum yx_1 = b_0\sum x_1 + b_1\sum x_1x_1 + b_2\sum x_1x_2 \,, \\ \sum yx_2 = b_0\sum x_1 + b_1\sum x_1x_2 + b_2\sum x_2x_2 \,. \end{cases}$$

Единственное отличие новой системы – первом уравнении отсутствует член nb_0.
С последних двух суммарных выражений можно получить:

$$\sum y^2 = b_1\sum yx_1 + b_2\sum yx_2 + b_1\sum ex_1 + b_2\sum ex_2 + \sum e^2 =$$
$$= b_1\sum yx_1 + b_2\sum yx_2 + \lambda(b_1\sum x_1 + b_2\sum x_2) + \sum e^2 =$$
$$=\lambda\sum y + b_1\sum yx_1 + b_2\sum yx_2 + \sum e^2,$$

что очень напоминает аналогичное выражение для модели со свободным членом (с заменой λ на b_0).

С истории науки изчестно, что в 1927 г. Экономист Пол Дуглас, рассматривая диаграммы логарифмов капитальных затрат (K), объемов выпуска продукции (Y) и затрат труда (L), выявил, что расстояния от точек графика ln$Y$ до точек графиков ln$K$, ln$L$ составляют одинаковую пропорцию для всх наблюдений с 1900 по 1922 гг. На основании этого факта математик Чарльз Кобб показал, что такая особенность имеет место для зависимости $Y = K^bL^{1-b}$, имеющей сейчас название Кобба – Дугласа.

Свое открытие П. Дуглас сделал лишь потому, что для 1989 года он принял значения всех показателей K, Y, L за 1 (100%), то есть задав фиксированный узел, где пересекаются графики зависимостей ln$K$, ln$Y$, ln$L$ по времени ($x = t - 1989$).

Фактически он увидел линейность трендов логарифмов показателей (рис. 9.6) и этого факта было достаточно, чтобы сделать такие выводы.

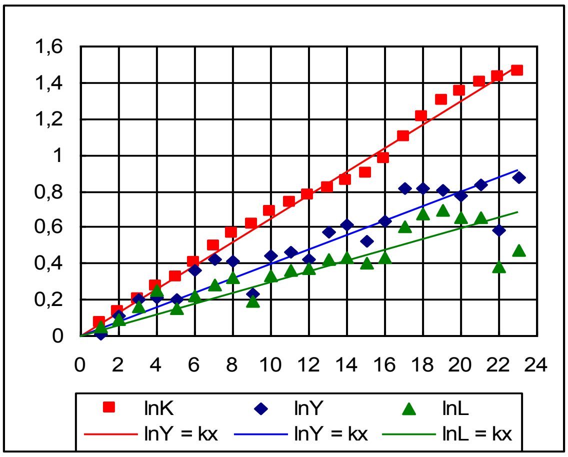


Рис. 9.6. Наблюдения П. Дугласа

Рассмотрим детально аппроксимацию $lnL = f(x)$. Данные П. Дугласа представлены в табл. 9.6. Если аппроксимировать зависимость обычной линейной моделью

$$lnL = b_0 + b_1 x + e,$$

то будут получены такие МНК-оценки параметров $b_0 = 0{,}08299$; $b_1 = 0{,}02435$.

Сумма остатков модели равна нулю $\sum e = 0$, сумма квадратов остатков $\sum e^2 = 0{,}1829$. Выполняется соотношение $s_y^2 = s_p^2 + s_e^2$, где обозначено $y = lnL$,

$y_p = b_0 + b_1 x$, s_p^2 – дисперсия вычисленных значений. В числах имеем разложение общей дисперсии на две составляющие 0,03404 = 0,02608 + 0,00796, откуда находим коэффициент детерминации $R^2 = 0,7664$. Через фиксированный узел линия регрессии не проходит (рис. 3).

Таблица 9.6

Исходные данные *K, Y, L*

№	*t*	*K*	*Y*	*L*	*lnK*	*lnY*	*lnL*
1	1900	1,07	1,01	1,05	0,0677	0,0100	0,0488
2	1901	1,14	1,12	1,10	0,1310	0,1133	0,0953
3	1902	1,22	1,22	1,18	0,1989	0,1989	0,1655
4	1903	1,31	1,24	1,29	0,2700	0,2151	0,2546
5	1904	1,38	1,22	1,16	0,3221	0,1989	0,1484
6	1905	1,49	1,43	1,25	0,3988	0,3577	0,2231
7	1906	1,63	1,52	1,33	0,4886	0,4187	0,2852
8	1907	1,76	1,51	1,38	0,5653	0,4121	0,3221
9	1908	1,85	1,26	1,21	0,6152	0,2311	0,1906
10	1909	1,98	1,55	1,40	0,6831	0,4383	0,3365
11	1910	2,08	1,59	1,44	0,7324	0,4637	0,3646
12	1911	2,16	1,53	1,45	0,7701	0,4253	0,3716
13	1912	2,26	1,77	1,52	0,8154	0,5710	0,4187
14	1913	2,36	1,84	1,54	0,8587	0,6098	0,4318
15	1914	2,44	1,69	1,49	0,8920	0,5247	0,3988
16	1915	2,66	1,89	1,54	0,9783	0,6366	0,4318
17	1916	2,98	2,25	1,82	1,0919	0,8109	0,5988
18	1917	3,35	2,27	1,96	1,2090	0,8198	0,6729
19	1918	3,66	2,23	2,00	1,2975	0,8020	0,6931
20	1919	3,87	2,18	1,93	1,3533	0,7793	0,6575
21	1920	4,07	2,31	1,93	1,4036	0,8372	0,6575
22	1921	4,17	1,79	1,47	1,4279	0,5822	0,3853
23	1922	4,31	2,40	1,61	1,4609	0,8755	0,4762

Если мы примем модель без свободного члена $lnL = kx + e$, то получим одержимо $k = 0,02965$. График зависимости проходит через фиксированный узел (рис. 3). Сумма остатков этой модели не равна нулю $\sum e = 0,4467$, сумма квадратов остатков $\sum e^2 = 0,2200$ (это больше сравнительно с предыдущей моделью. Не выполняется соотношение $s_y^2 = s_p^2 + s_e^2$, поэтому нельзя вычислить коэффициент детерминации, показывающий часть изменчивости, которая объясняется моделью. Также нельзя вычислить другие статистические характеристики.

Оценим параметры модели без свободного члена другим методом. Составим систему уравнений для определения k и λ:

$$\sum y = k\sum x,$$
$$\sum yx = \lambda\sum x + k\sum x^2,$$

$$\sum y^2 = \lambda\sum y + k\sum yx + \sum e^2$$

С первого уравнения этой системы, которое эквивалентно условию $\sum e = 0$, получим новую оценку параметра k:

$$k = \frac{\sum y}{\sum x} = \frac{8{,}6289}{276} = 0{,}03126.$$

Для однофакторной модели по данному методу предлагается определять единственный параметр модели k с обязательным условием $\sum e = 0$, не с $\sum e^2 \to min$. Со второго уравнения дополнительно находим $\lambda = -0{,}02536$. Третье соотношение можно использовать для контроля всех вычислений. Сумма остатков этой модели равна нулю $\sum e = 0$, сумма квадратов остатков $\sum e^2 = 0{,}2313$ (это больше сравнительно с предыдущими моделями). Соотношение $s_y^2 = s_p^2 + s_e^2$ не выполняется. Графики всех трех моделей представлены на рис. 9.7.

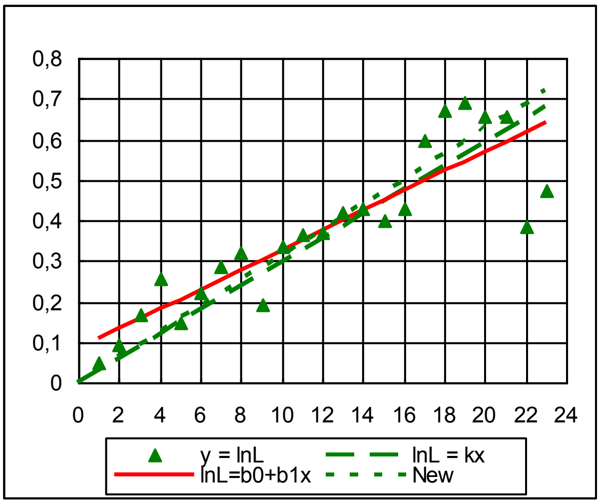


Рис. 9.7. Аппроксимация временного тренда для *lnL*

Следовательно, данный метод не имеет особенных преимуществ сравнительно со стандартным МНК, единственно, что достигнуто – это выполнение условий равенства нулю суммы остатков модели.

Вопросы для самопроверки:

1) Для каких целей рекомендуется дополнительно к «чистой регрессии» строить регрессионную модель с β -коэффициентами?

2) В каких пределах изменяются β -коэффициенты?

3) Могут ли в экономике встречаться регрессионные модели в стандартизированных переменных с $|\beta| \geq 1$?

4) Какие причины обуславливают $|\beta| \geq 1$?

5) Какие существуют рекомендации при построении линейной регрессионной модели, чтобы $|\beta|<1$?

6) Приведите пример классических эконометрических моделей в экономике без свободного члена.

7) Через какой узел проходит линия регрессии без свободного члена?

8) Какие математические проблемы существуют в регрессионной модели без свободного члена?

9) Какие существуют рекомендации в решении математических проблем в регрессионной модели без свободного члена?

Тема 10. Обобщенные схемы регрессионного анализа

10.1. Некоторые альтернативные схемы регрессионного анализа

10.2. Модели с Dummy-переменными

10.3. Новейшие (Advanced) методы регрессионного анализа

10.1. Некоторые альтернативные схемы регрессионного анализа

Совершенствование регрессионного анализа, как методики оценки параметров моделей, пошло по двум направлениям. Во-первых, усложняется стандартная процедура МНК так, чтобы она была в состоянии преодолевать те или иные затруднения (метод пошаговой регрессии, метод псевдообратной матрицы, метод SVD, двухшаговая и трехшаговая процедура МНК, обобщенный метод НК). Во-вторых, делаются попытки вообще отказаться от МНК и заменить его на совершенно другую процедуру оценки параметров. В некоторых предложениях вносится столько изменений и дополнений в стандартный МНК, что, по существу, такие предложения следует также рассматривать как новые (альтернативные) схемы регрессионного анализа.

Метод ортогональной (нормальной) регрессии

Название метода не имеет ничего общего ни с ортогонализацией переменных (подобно методу ортогонализации Грамма-Шмидта или методу главных компонент), ни с классическим определением “регрессионной (корреляционной) зависимости”.

Согласно первой предпосылке регрессионного анализа, в стандартном МНК необходимо одну из переменных обязательно объявить результативным признаком и все случайные ошибки приписать только этой переменной (ошибки измеряются вдоль оси избранной переменной). В результате получается одна из нескольких (совершенно различных) сопряженных регрессионных моделей. Эта первая предпосылка МНК неприемлема в случае равноправия всех переменных, когда совместная их изменчивость определяется тем, что они являются следствиями одной и той же причины. В методе "*ортогональной регрессии*" предлагается случайную ошибку измерять ортогонально (по нормали) к поверхности регрессии, а не вдоль какой-либо координатной оси.

Все дальнейшие рассуждения проводим в пространстве стандартизованных переменных, так как при изменении масштабов переменных условие ортогональности нарушается.

Чтобы найти расстояние ε от гиперплоскости

$$a_0 x_0 + a_1 x_1 + a_2 x_2 + ... + a_m x_m = 0$$

до произвольной точки, надо нормировать все коэффициенты (привести уравнение плоскости к нормальной форме) и подставить в это уравнение координаты интересующей нас точки:

.

Здесь переменная y обозначена через x_0, чтобы не выделять ее.

Мы получили задачу на условный экстремум - необходимо найти минимум суммы квадратов ошибок [ε^2] при условии нормирования коэффициентов .

Преобразуем сумму квадратов ошибок, учитывая, что для стандартизованных переменных :

.

Составляем функцию Лагранжа

и приравниваем нулю ее частные производные по a_i . Получаем задачу на собственные значения и собственные векторы (расширенной) корреляционной матрицы (включая корреляции с y):

.

Таким образом получили коэффициенты $a_{i.}$

Гребневая регрессия

Оценки МНК параметров модели являются состоятельными, несмещенными и эффективными; им соответствует минимальное значение суммы квадратов ошибок $[e^2]$. Все это хорошо, но при наличии мультиколинеарности МНК-оценки становятся неустойчивыми. Предложено (Гоэрл) перед обращением матрицы X^TX добавить к ее диагональным элементам малые числа k, что предотвращает вырождение матрицы, увеличивает ее определитель и уменьшает дисперсии параметров.

Новые оценки (ридж-оценки) параметров вычисляются по формулам

$$\boldsymbol{b} = (X^T X + kI)^{-1} X^T Y\,.$$

От такой операции (увеличения диагональных элементов матрицы X^TX на малое число k) сумма квадратов ошибок увеличивается, но типичный характер зависимости остаточной дисперсии от k подобен графику на рис. 1 - до какой-то границы $k < k^*$ остаточная дисперсия возрастает незначительно.

Рис. 1. Возрастание остаточной дисперсии при увеличении диагональных элементов матрицы X^TX на k

Устойчивость оценок параметров характеризуется суммой их дисперсий

.

Эту формулу можно преобразовать к более ясному виду, если найти все собственные числа λ_i и собственные векторы U_i матрицы $X^TX = UD_\lambda U^T$. Здесь через D_λ обозначена диагональная матрица собственных чисел, а через U - ортогональная матрица собственных векторов.

Вычисляем $U^T \boldsymbol{b} = \boldsymbol{g}$.

Теперь сумма дисперсий параметров модели может быть представлена в форме

.

При мультиколлинеарности (вырождении матрицы X^TX) одно или несколько собственных чисел λ_i близки к нулю и при $k=0$ получается очень большое (если не бесконечное) значение $L(\boldsymbol{b})$. Введение даже очень малого k сразу исправляет положение.

Существует несколько рекомендаций по выбору величины k. Выбирают такое наименьшее значение параметра k, при котором поведение всех коэффициентов регрессии уже стабилизировалось, но еще не начался интенсивный рост остаточной дисперсии.

Для сравнения различных альтернативных схем регрессионного анализа были предприняты многочисленные проверки на данных с различными особенностями. Первое место во всех таких проверках уверенно заняли ридж-оценки. Некоторые методы были специально разработаны для преодоления конкретных особенностей; в этих случаях ридж-оценки делили первое-второе место с одним из специализированных методов. МНК-оценки во всех проверках заняли последнее место по всем критериям качества. Общее заключение - если мультиколинеарности нет, то ридж-оценки близки к МНК-оценкам, а в случае мультиколинеарности они значительно лучше, причем преимущество ридж-оценок увеличивается с ростом числа переменных m.

10.2. Модели с *dummy*-переменными

Dummy в переводе с английского языка обозначает «макет», поэтому иногда встречается название этих переменных «макетные переменные», а иногда и неправильное название «фиктивные переменные».

Dummy-переменные предназначены для учета в модели качественных переменных, измеренных на номинальных шкалах. Например, пол: женский, мужской; состояние экономики: до кризиса и после; сезон: зимний, весенний, летний, осенний (это категории качественной переменной). Если в модель необходимо ввести переменную, имеющую несколько категорий, вводится соответствующее количество *dummy*-переменных, она равна единице, если наблюдается соответствующая категория и нулю, если она отсутствует в наблюдении. В сумме все dummy-переменные образуют $X_0 \equiv 1$, поэтому нельзя включать в модель сразу все *dummy*-переменные или же не надо включать в модель свободный член. Конечно, одну из категорий берут эталонной и ее в модель не включают. Поэтому получаем модель для эталонной категории с исправлением для всех других категорий. Значимость этих исправлений по критерию Стьюдента есть значимостью отличий каждой категории от эталонной.

Пример. Известны затраты на газ и электроэнергию в США в течении 1977 – 1982 г.г. Имеем типичную ситуацию, когда желая увеличить объем выборки, вместо среднегодовых используют квартальные данные. Не смотря, что выборка при этом увеличилась в четыре раза, значимость модели резко ухудшилась, в коэффициент детерминации уменьшился до нуля. Причина заключается в том, что при переходе к квартальным данным появились сезонные колебания, на интенсивном фоне которых совсем потерялся исследуемый эффект – временной линейный тренд (рис. 10.1).

Следовательно, необходимо учесть в модели эту особенность процесса, и поэтому самым экономным способом является введение *dummy*-переменныех для описания межквартальных отличий. Поэтому необходимо ввести в модель четыре дополнительные *dummy*-переменные: $Z_1 = 1$ – для 1-го квартала и $Z_1 = 0$ – для всех остальных кварталов.; $Z_2 = 1$ – для 2-го квартала и $Z_2 = 0$ – для всех остальных кварталов; $Z_3 = 1$ – для 3-го квартала и $Z_3 = 0$ – для всех остальных кварталов; $Z_4 = 1$ – для 4-го квартала и $Z_4 = 0$ – для всех остальных кварталов. Если за эталон взять 1-й квартал, то переменные Z_1 не следует включать в модель. В результате вычислений будет получено уравнение регрессии для эталонной категории с исправлениями для свободного члена для других категорий.

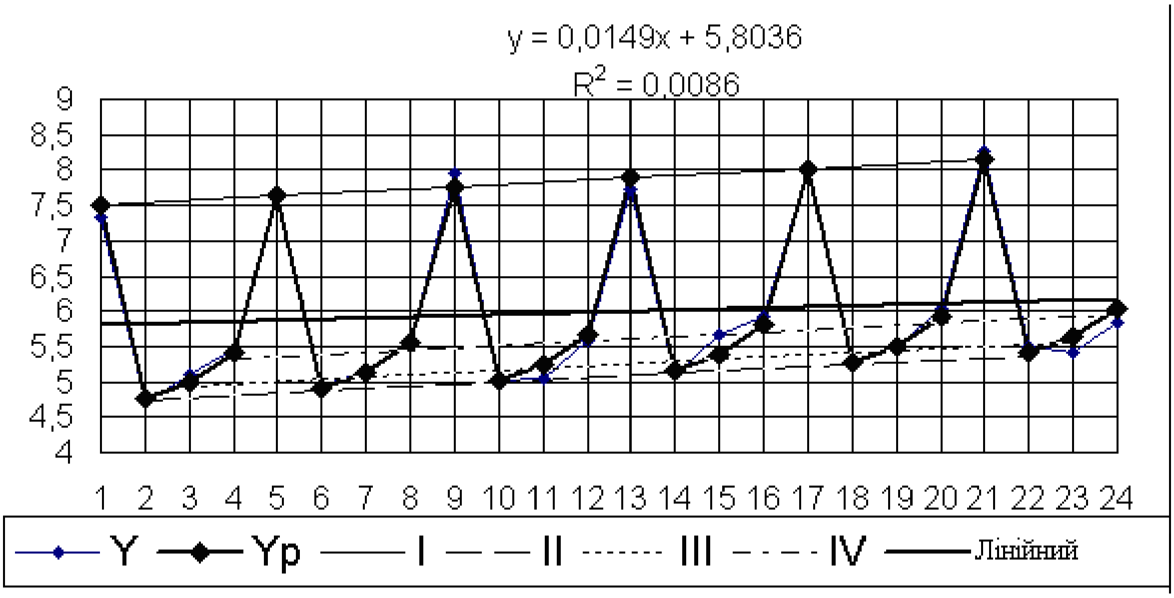


Рис. 10.1. Сезонные колебания данных по кварталам

Можно подключить в модель все сразу *dummy*-переменные, но тогда в модель следует включить свободный член, поскольку $Z_1 + Z_2 + Z_3 + Z_4 = 1$. В результате таких вычислений будут получены уравнения регрессии для каждой категории отдельно и правильно подсчитаны все статистические характеристики за исключением статистики Фишера, которая будет занижена в $\frac{m+1}{m} = \frac{5}{4}$. При этих разных подходах по критерию Стьюдента оцениваются разные эффекты качественного фактора. По другой методике оценивается значимость эффекта каждой категории, а по первой, рекомендуемой, значимость расхождений каждой категории от эталонной. Сейчас неявно предусматривается, что для каждой совокупности (для каждой категории) сохраняются неизменными все закономерности в зависимостях от количественных переменных, а влияние качественного признака проявляется только в управлениях свободного члена. Бывают и более сложные ситуации, когда для разных категорий выявляются разные эффекты количественных факторов. Тогда в модели следует учесть также члены со взаимосвязями. Исходные данные вместе с введенными *dummy*-переменными представлены в табл. 10.1, где T -номер квартала, Y - затраты (результативный признак).

Таблица 10.1

Данные для построения модели

Z_1	Z_2	Z_3	Z_4	T	Y	Y_p
1	0	0	0	1	7,33	7,5141
0	1	0	0	2	4,7	4,7691
0	0	1	0	3	5,1	4,9991
0	0	0	1	4	5,46	5,4158
1	0	0	0	5	7,65	7,6405
0	1	0	0	6	4,92	4,8955
0	0	1	0	7	5,15	5,1255
0	0	0	1	8	5,55	5,5421
1	0	0	0	9	7,96	7,7668
0	1	0	0	10	5,01	5,0218
0	0	1	0	11	5,05	5,2518
0	0	0	1	12	5,59	5,6685
1	0	0	0	13	7,74	7,8932
0	1	0	0	14	5,1	5,1482
0	0	1	0	15	5,67	5,3782
0	0	0	1	16	5,92	5,7948
1	0	0	0	17	8,04	8,0195
0	1	0	0	18	5,27	5,2745
0	0	1	0	19	5,51	5,5045
0	0	0	1	20	6,04	5,9212
1	0	0	0	21	8,26	8,1459
0	1	0	0	22	5,51	5,4009
0	0	1	0	23	5,41	5,6309
0	0	0	1	24	5,83	6,0476

Значимость коэффициентов регрессии перед Z_2, Z_3, Z_4 в этой модели показывают, что затраты на газ и электроэнергию в эталонном 1-м квартале существенно отличаются от других кварталов. Приведем уравнения регрессии для каждого квартала:

Yp = 7,4825 – 2,777· Z2 – 2,278· Z3 – 2,193· Z4 + 0,0316· T; R^2 = 0,9867
(tb) (98,5) (33,2) (30,7) (25,9) (7,3)

$Y_p(\mathrm{I}) = 7{,}4825 + 0{,}0316 \cdot T$ ($Z_1 = 1$),

$Y_p(\mathrm{II}) = 7{,}4825 – 2{,}777 + 0{,}0316 \cdot T = 4{,}7059 + 0{,}0316 \cdot T$ ($Z_2 = 1$),

$Y_p(\mathrm{III}) = 7{,}4825 – 2{,}278 + 0{,}0316 \cdot T = 4{,}9034 + 0{,}0316 \cdot T$ ($Z_3 = 1$),

$Y_p(\mathrm{IV}) = 7{,}4825 – 2{,}193 + 0{,}0316 \cdot T = 5{,}2894 + 0{,}0316 \cdot T$ ($Z_4 = 1$).

На фоне интенсивных сезонных колебаний значимость исследуемого линейного тренда оказалась заниженной до нуля ($R^2 = 0{,}0086$), оценка углового коэффициента занижена более чем в два раза.

С помощью трех дополнительных *dummy*-переменных Z_2, Z_3, Z_4 очень хорошо описана вся зависимость с сезонными колебаниями. Отметим, что попытки описать эти колебания аналитическим выражением в виде трех гармоник (с такими периодами, как год, полгода, квартал) будут неэкономными и потребуют 6-ти дополнительных параметров (по два параметра на каждую гармонику). Можно в модель включить все сразу *dummy*-переменные, но тогда модель не должна иметь свободный член. Вычисляем параметры такой линейной модели:

$$Y_p = (b_1Z_1 + b_2Z_2 + b_3Z_3 + b_4Z_4) + b_5T \quad \text{(константа отсутствует).}$$

В таком случае будут получены все уравнения для каждого квартала (не надо пересчитывать свободный член). Однако теперь статистики Стьюдента будут оценивать отклонения среднего уровня затрат в квартале от нуля, а не от выбранного эталона. Все характеристики, кроме F, получатся такие же, только статистику Фишера следует увеличить в $(m + 1) / m = 5 / 4$ раз: $F = 280{,}93 \cdot 5/4 = 351{,}16$.

10.3. Новейшие (Advanced) методы регрессионного анализа

В новейших методах регрессионного анализа результативный признак может быть дискретным и даже качественным, выраженным категориями или номинациями. В этих методах изучается вероятность этих категорий в зависимости от значений объясняющих переменных.

В общем регрессионном анализе на основе внешнего поведения результативного признака, заданного некоторым количеством случайных наблюдений, определяют форму связи и далее методом наименьших квадратов оценивают ее параметры или коэффициенты регрессии. В линейном регрессионном анализе форма связи принимается линейной относительно параметров модели, другими словами, возможны предварительные преобразования переменных. В результате этого, получается, что результативный признак всегда будет распределен асимптотически нормально со своими переменными характеристиками (средним и дисперсией) для каждого набора объясняющих переменных.

В обобщенной линейной модели вид закона распределения результативного признака принимается на основе содержательных обоснований, выполняется аппроксимация не конечной зависимости $y(x)$, а некоторого параметра функции распределения. Далее, зная распределение результативного признака, для каждого набора объясняющих переменных вычисляют среднее (ожидаемое) значение результативного признака и ее дисперсию. Считается, что один из параметров функции распределения, а именно линейный предиктор, зависит от объясняющих переменных и принимается его аппроксимация в виде линейной комбинации объясняющих переменных, линейной относительно ее коэффициентов, которые оцениваются методом наименьших квадратов. Так, для описания зависимости от одной объясняющей переменной в традиционном регрессионном анализе принимают линейную модель, имеющую вид: $\hat{y}(x) = b_0 + b_1 x$, где $\hat{y}(x) = y_p(x)$ – ожидаемое (среднее) значение *y* для каждого *x*; возможны предварительные функциональные преобразования начальных переменных (X, Y). Коэффициенты регрессии a, b оценивают методом наименьших квадратов, после чего определяют несмещенную оценку остаточной $s_e^2 = Const$ и дисперсии вычисленных значений для каждого *x*: $\sigma^2(\hat{y}) = \frac{s_e^2}{N}\left(1 + \frac{(x-\bar{x})^2}{s_x^2}\right)$, где N – число наблюдений, s_x^2 – дисперсия объясняющей переменной *x*. Границы 95 %-го доверительного интервала на каждое вычисленное значение $\hat{y}(x)$ вычисляется по формулам:

$$y_d = \hat{y} - t_{0,05}(N-2)\cdot\sigma(\hat{y}); \quad y_u = \hat{y} + t_{0,05}(N-2)\cdot\sigma(\hat{y}),$$

где y_d – нижняя граница (*down*), y_u – верхняя граница (*up*), $t_{0,05}(dfe)$ – квантиль распределения Стьюдента, $dfe = (N-2)$ – остаточное число степеней свободы. В обобщенной линейной модели считается, что результативный признак Y является случайной величиной с известным видом закона распределения, который описывается функцией плотности вероятности $f(y \mid \mu, \nu)$, где μ и ν – параметры функции распределения (вероятности – для дискретной или плотность вероятности – для непрерывной случайной величины). Принимается, что один из этих параметров, а именно линейный предиктор μ, зависит от объясняющих переменных, другой параметр, если закон распределения двухпараметрический, от объясняющих переменных не зависит – $\nu = Const$. Линейный предиктор является комбинацией объясняющих переменных и эта комбинация линейна относительно параметров: $\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m$ (допускаются также предварительные функциональные преобразования начальных переменных). По экспериментальным данным (x_i, y_i) методом наименьших квадратов оцениваются коэффициенты линейного предиктора β_j и константа ν. Предполагается, что функциональная зависимость между предиктором и ожидаемыми значениями результативной переменной $\mu = g(M(y)) = g(\hat{y})$. Функция g считается известной и называется функцией связи. Традиционная модель регрессионного анализа является отдельным случаем обобщенной линейной модели при тождественной форме связи $\hat{y} = \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m$ и нормальном законе распределения $Y \sim N(\mu, \sigma)$. Укажем, что даже в этом случаи регрессионная модель и обобщенная линейная модель не эквивалентны, поскольку в них разные методы определения ширины доверительной полосы на вычисленные значения.

Биномиальная модель («логит-» и «пробит-» анализы).

Пример. Имеем исследование влияния затрат на рекламу на объемы реализации мобильных телефонов. В табл. 10.2 приведены данные об относительной части реализованных телефонов в филиалах фирмы, имеющих разные затраты на рекламу. Количество реализованных телефонов m в партиях товара с n единиц для каждого уровня затрат на рекламу x распределено по биномиальному закону $P(Y = m) = P_n(m) = C_n^m p^m (1-p)^{n-m}$ с математическим ожиданием (средним) $y_p = M(n) = np$ и дисперсией $D(n) = np(1-p)$.

Вероятность $P_n(m)$ зависит от двух параметров: n – количества товара в партии и $p(x)$ – удельной относительной части реализованных телефонов. Относительная часть реализованных телефонов $0 \le p \le 1$ будет большой (близкой к 1) при больших затратах на рекламу и маленькой (близкой к нулю) при небольших.

Таблица 10.2

Затраты на рекламу в филиалах

Филиал	Затраты на рекламу (x)	Объем партии товара (n)	Количество проданных телефонов (m)	Относительная часть реализованных телефонов (p)
1	10,2	50	44	0,8800
2	7,7	49	42	0,8571
3	5,1	46	24	0,5217
4	3,8	48	16	0,3333
5	2,6	50	6	0,1200

Необходимо аппроксимировать зависимость $p(x)$ достаточно простым выражением. По данным табл. 10. 2 построен график p от x (рис. 2). Экспериментальных точек мало, хотя каждая точка является результатом $n \approx 50$ наблюдений. По очевидным рассуждениям ясно, что эта сигмоидальная зависимость должна быть похожей на интегральную функцию распределения (не обязательно нормального закона) или на логистическую зависимость. Забегая наперед отметим, что обе аппроксимации привели к практически одинаковым результатам (на рис. 10.2 сплошная и пунктирная линии).

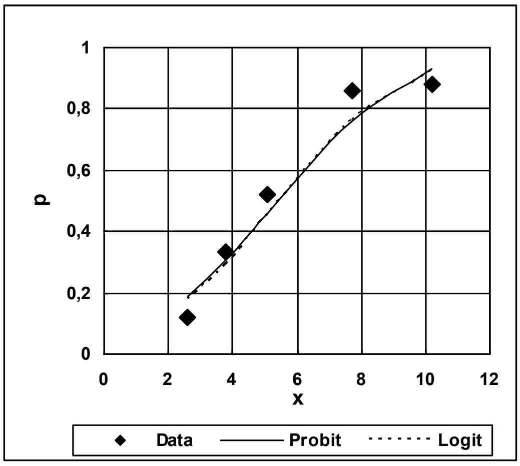


Рис. 10.2. Аппроксимация относительной части реализованного товара в зависимости то затрат на рекламу

Если применяется аппроксимация функцией нормального распределения, используется название «пробит-анализ»; если применяется аппроксимация логистической функцией, используется название «логит-анализ». Сравним оба вида биномиальной модели.

Пробит-анализ. Аппроксимируем зависимость параметра $p(x)$ биномиального распределения интегральной функцией стандартного нормального закона:

$$p(x) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{z^2}{2}} dz \quad \leftrightarrow \quad t = \Phi^{-1}(p),$$

где Φ^{-1} – функция, обратная к функции распределения нормального закона, а $t = \dfrac{x-a}{\sigma} = b_0 + b_1 x$ – линейный предиктор (откуда ис следует название «линейная» модель).

Для каждого известного значения p с табл. 10.2 находим квантили стандартного нормального распределения $t = \Phi^{-1}(p)$ (табл. 10.3) и обозначаем их в виде точек (x–t) на рис. 3 *а*. Далее методом наименьших квадратов оцениваем коэффициенты регрессии линейного предиктора $t = 0{,}3113x - 1{,}6922$ и в табл. 3 вычисляем расчетные (сглаженные) значения:

$$p_n = \Phi(0{,}3113x - 1{,}6822).$$

Таблица 10.3

Расчетные значения $p(x)$ в пробит- (p_n) и логит- анализе (p_s)

№	*x*	*p*	*t*	p_n	*s*	p_s
1	10,2	0,8800	1,174987	0,930978	1,99243	0,924514
2	7,7	0,8571	1,067571	0,759548	1,81529	0,767224
3	5,1	0,5217	0,054519	0,458368	0,080043	0,457004
4	3,8	0,3333	-0,43073	0,305291	-0,70819	0,298392
5	2,6	0,1200	-1,17499	0,188669	-1,99243	0,184671

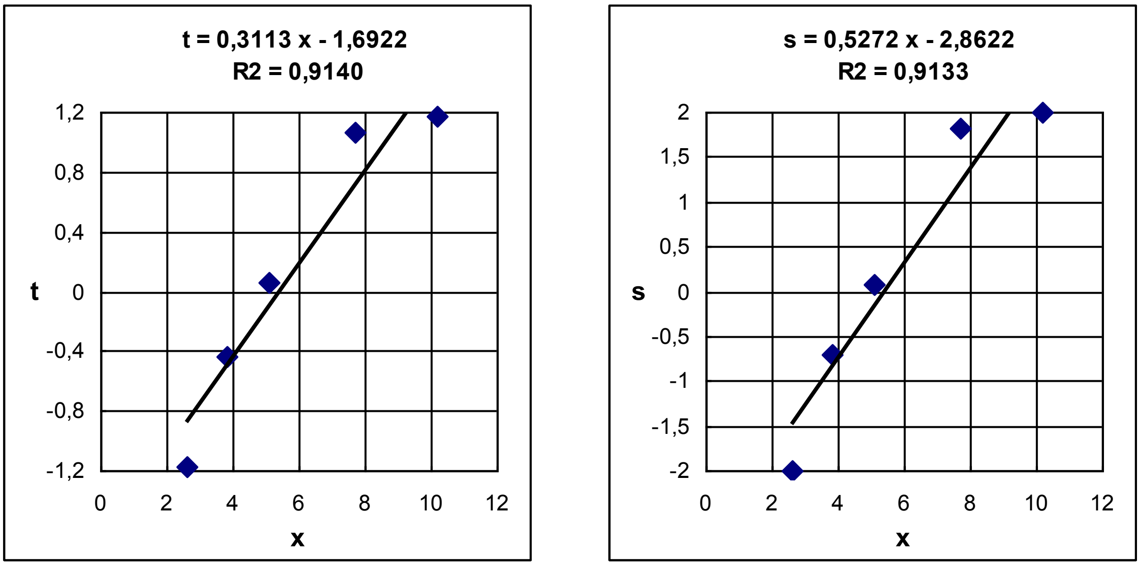
Для сравнения сразу приведем вычисления по логит-анализа.

Логит-анализ. Аппроксимируем зависимость параметра $p(x)$ биномиального распределения логистической функцией:

$$p(x) = \frac{1}{1+e^{-s}} = \frac{1}{1+e^{-(a+bx)}} \quad \leftrightarrow \quad s = \ln\frac{p}{1-p},$$

где $s = a + bx$ – линейный предиктор.

Для кожного известного значения p с табл. 2 находим значение $s = ln\dfrac{p}{1-p}$ (табл. 3) и отмечаем их в виде точек (x–s) на графике рис. 3 - *б*. Далее методом наименьших квадратов оцениваем коэффициенты регрессии линейного предиктора $s = 0{,}5272x - 2{,}8622$ и в табл. 10.3 вычисляем расчетные (сглаженные) значения $p_s = \dfrac{1}{1 + e^{-(0{,}5272\,x - 2{,}8622)}}$.

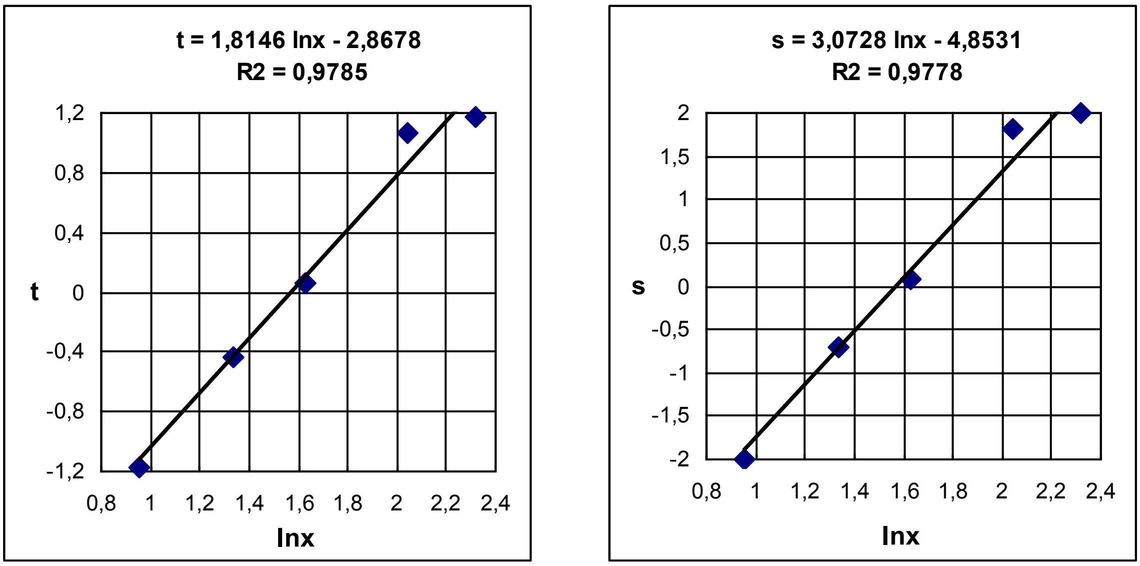


а) б)

Рис. 10.3. Аппроксимация зависимости $p(x)$ в функциональных переменных
а) пробит-анализ: $t = \Phi^{-1}(p(x))$; б) логит-анализ: $s = p(x) / (1 - p(x))$

На рис. 2 построены графики $p_n(x)$ – сплошной линией, $p_s(x)$ – пунктирной линией. Видим, что вычисленные значения p_n, p_s по двум видам анализа (пробит- и логит-) близки между собой, то есть оба вида анализа эквивалентны. Это не случайно, поскольку логистической зависимостью можно хорошо аппроксимировать функцию распределения нормального закона. Как определенное преимущество логит-анализа отметим более простой вид математических формул, записанных с помощью элементарных функций.

Если правильно принято предположение о логистической форме связи, точки на графике (x–s) должны группироваться около некоторой прямой, однако здесь (как и для пробит-анализа), есть явные отклонения от прямой. В обобщенных линейных моделях предполагаются предварительные функциональные преобразования переменных. Так в приведенном примере целесообразно перейти к логарифмам объясняющей переменной. При этом существенно улучшиться качество аппроксимации как при пробит-, так и при логит-анализе (рис. 10.4).



а) б)

Рис. 10.4. Аппроксимация зависимости $p(\ln x)$ в функциональных переменных а) пробит-анализ: $t = \Phi^{-1}(p(\ln x))$; б) логит-анализ: $s = \ln(p(\ln x)/(1 - p(\ln x)))$

С рис. 10.4 видно, что коэффициент детерминации после перехода от логарифмов увеличился от 0,91 до 0,98. Теперь можно назвать обобщенную линейную модель обобщенной логарифмической моделью. Следовательно, функция распределения результативной переменной (y – количество реализованных телефонов) определена полностью. Вона зависит от двух параметров $n = const$ и $p = f(x)$:

$$p = \frac{1}{1+e^{-s}} = \frac{1}{1+e^{-(3{,}0728 \ln x - 4{,}8531)}} \approx \frac{x^3}{x^3+100}.$$

Для известного n = 50 та для каждого x = 0, 1, 2, 3 ..., 12 в табл. 10.4 вычислены средние значения (математические ожидания) результативной переменной $Y = np(x)$ и средние квадратические отклонения $\sigma = \sqrt{np(x)(1 - p(x))}$.

При $n = 50$ биномиальное распределение уже асимптотически нормальное (распределение Лапласа), поэтому границы 95 %-х доверительных интервалов определяем по правилу: $Y_d = Y - t_{0,05} \cdot \sigma$ – нижняя граница (*down*), $y_u = Y + t_{0,05} \cdot \sigma$ – верхняя (*up*) межа, $t_{0,05}$ = 2,01 находим с таблиц Стьюдента при $dfe = n - 2 = 48$.

На рис. 10.5 изображен график зависимости $y(x)$ вместе с 95 %-й доверительной полосой. Точками обозначены наблюдаемые значения m (скорректированные на одинаковый объем партий n = 50).

Таблица 10.4

Вычисленные значения y с 95 %-ми доверительными полосами

x	s	p	σ	$Y = np$	-95%	+95%
1	-4,853	0,008	0,620	0,387	-0,859	1,633
2	-2,723	0,062	1,700	3,081	-0,337	6,499
3	-1,477	0,186	2,750	9,292	3,763	14,820
4	-0,593	0,356	3,385	17,794	10,989	24,599
5	0,092	0,523	3,532	26,154	19,055	33,253
6	0,653	0,658	3,355	32,880	26,136	39,624
7	1,126	0,755	3,041	37,758	31,646	43,869
8	1,537	0,823	2,699	41,149	35,724	46,574
9	1,899	0,870	2,380	43,486	38,702	48,270
10	2,222	0,902	2,100	45,112	40,890	49,333
11	2,515	0,925	1,860	46,260	42,521	49,999
12	2,783	0,942	1,657	47,086	43,757	50,416

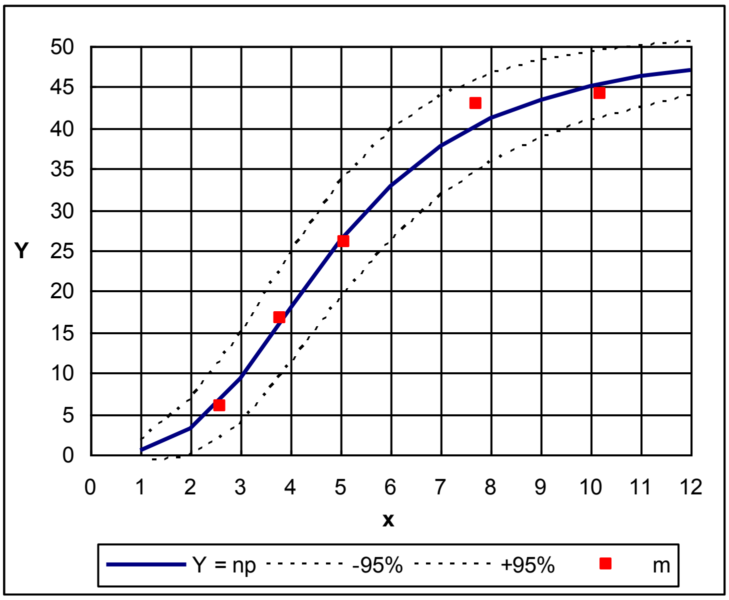


Рис. 10.5. Результаты анализа по обобщенной модели

На этом же примере сравним результаты проведенных вычислений (по обобщенной линейной модели) с результатами вычислений по нелинейному регрессионному анализе. Пусть относительная часть $p(x)$ является результативным признаком. Используя предыдущее исследование, можно получить ту же самую форму связи для описания выбранного результативного признака:

$$y = p = \frac{1}{1+e^{-s(x)}} \leftrightarrow s(x) = \ln\frac{p}{1-p},$$

$$\hat{s} = 3{,}0728 \cdot \ln x - 4{,}8531.$$

У функциональных координатах ($X = lnx,\ s$) изучаемая зависимость является линейной. Следовательно, до этого момента методики обобщенной линейной модели и регрессионного анализа совпадали (естественно, только для данного простого примера). Однако, далее начинаются существенные расхождения.

В обобщенной линейной модели $p(x)$ считается параметром биномиального распределения результативного признака (второй параметр $n = 50$). Зная функцию распределения, для каждого значения аргумента x вычисляем ожидаемое значение $\hat{y} = M\left(\frac{m}{n}\right) = \frac{n\hat{p}}{n} = \hat{p}$ и дисперсию $\sigma^2(\hat{y}) = D\left(\frac{m}{n}\right) = \frac{n\hat{p}(1-\hat{p})}{n^2} = \frac{\hat{p}(1-\hat{p})}{n}$. График уравнения регрессии для $y = p(x)$ вместе с границами 95 %-го интервала (рис. 6 -а) только масштабом оси ординат отличается от графика на рис. 5.

В традиционном регрессионном анализе принята совсем другая методика определения границ доверительной полосы. Параметр $n = Const$ (число объектов в группе для вычисления относительной части $p = {}^{m}/_{n}$) вообще не входит в модель. Считается, что имеется всего $N = 5$ наблюдений (5 точек) в функциональных координатах (X, s). Для этих 5-ти точек находим несмещенную оценку дисперсии остатков линейной регрессионной модели $MSE = Const$, где $MSE = \frac{\sum (s-\hat{s})^2}{N-2} =$ 0,331079. Далее для каждого значения аргумента вычисляется дисперсия расчетных значений по формуле:

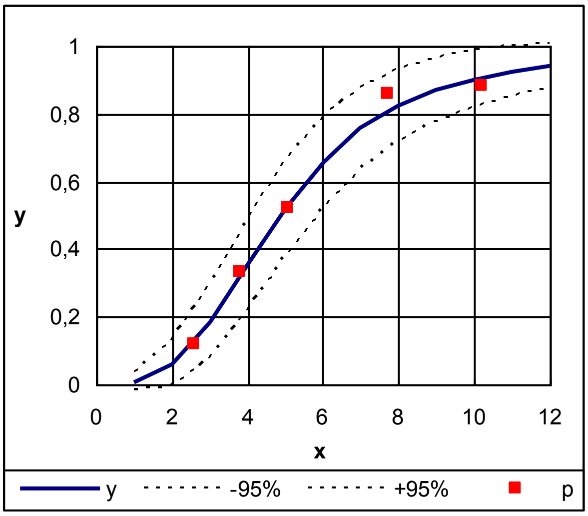
$$\sigma^2(\hat{s}) = \frac{MSE}{N}\left(1 + \frac{(X - \overline{X})^2}{s_X^2}\right) = \frac{0{,}331079}{5}\left(1 + \frac{(\ln x - 5{,}88)^2}{7{,}5336}\right).$$

В табл. 5 для x = 0, 1, 2 ..., 12 вычислены расчетные значения $\hat{s}$, их дисперсии $\sigma^2(\hat{s})$ и границы 95 %-й доверительной полосы $s_d,\ s_u$:

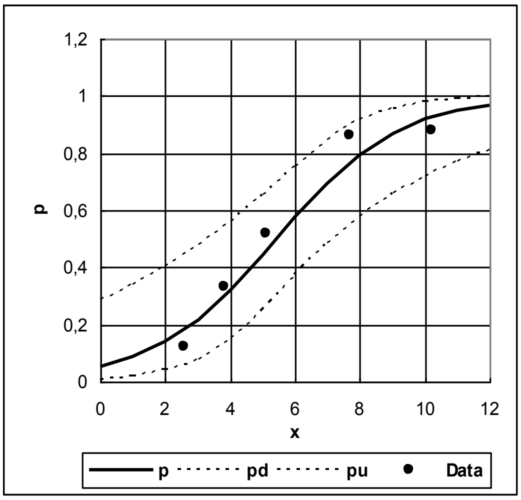
$$s_d = \hat{s} - t_{0,05}(N-2)\cdot\sigma(\hat{s}); \quad s_u = \hat{s} + t_{0,05}(N-2)\cdot\sigma(\hat{s});$$

где $t_{0,05}(3)$ = 3,1824 найдено по таблицам Стьюдента.

Далее выполнено обратное преобразование $p = \frac{1}{1+e^{-s}}$ и на основании этого построен график (рис. 10.6 -б). Сравнивая рис. 10.6. - а и рис. 10.6 -б отметим, что хотя линии регрессии одинаковые (для данного простого примера), но есть существенные отличия в ширине доверительной полосы (полосе неопределенности). Метод нелинейного регрессионного анализа привел к менее определенным результатам сравнительно с методикой обобщенной линейной модели.



а)



б)

Рис. 10.6. Сравнение обобщенной линейной модели (а) с нелинейным регрессионным анализом (б)

Таблица 5

Вычисление нелинейной регерссионной зависимости

X	s	σ^2	s_d	s_u
0	-2,862	0,370104	-4,79808	-0,92592
1	-2,3348	0,27553	-4,0053	-0,6643
2	-1,8076	0,198535	-3,22561	-0,38959
3	-1,2804	0,139119	-2,46741	-0,09339
4	-0,7532	0,097281	-1,7458	0,239402
5	-0,226	0,073022	-1,08598	0,633981
6	0,3012	0,066342	-0,5185	1,120903
7	0,8284	0,077241	-0,05608	1,712875
8	1,3556	0,105719	0,320845	2,390355
9	1,8828	0,151775	0,642971	3,122629
10	2,41	0,215411	0,932952	3,887048
11	2,9372	0,296625	1,203936	4,670464
12	3,4644	0,395417	1,463207	5,465593

Таким образом, обобщенные линейные модели предполагают более реалистическую постановку проблемы описания случайных зависимостей. Не надо больше визуально угадывать вид нелинейной регрессионной модели. Что не просто даже в однофакторных моделях. Выполняется аппроксимация не конечной нелинейной зависимости, а одного з параметров функции распределения, а именно линейного предиктора, то есть с вычислительной точки зрения остаемся в пределах линейных моделей. Вид функции распределения результативного признака считается известным заведомо с содержательных условий задания.

Вопросы для самопроверки:

1. В чем суть метода ортогональной (нормальной) регрессии?
2. Какие предназначения гребневой регрессии?
3. Какое предназначение *dummy*-переменных?
4. В каких случаях рекомендуется строить модели с *dummy*-переменными?
5. Какие особенности имеют новейшие методы регрессионного анализа?

6) В каких случаях рекомендуется строится биномиальная модель?

7. В чем идея пробит-анализа?

8) В чем идея логит-анализа?

Тема 11. Системы одновременных уравнений

11.1. Составляющие систем одновременных уравнений

11.2. Методы оценивания параметров систем уравнений. Косвенный метод наименьших квадратов (КМНК)

11.3. Проблема идентификации

11.4. Двухшаговый МНК (ДМНК)

11.1. Составляющие систем одновременных уравнений

Много сложных экономических процессов моделируются несколькими связанными между собой уравнениями, содержащими как повторяющиеся, так и собственные переменные. В этом случае говорят о системе одновременных уравнений. В одних уравнениях определенная переменная рассматривается как объясняющая (независимая), в другие уравнения системы она входит как зависимая переменная.

Одним из простейших примеров систем одновременных уравнений является модель «спрос-предложение», содержащая функции спроса, предложения линейные от цены и условие равновесия:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, \alpha_1 < 0, \\ q_t^S = \beta_0 + \beta_1 p_t + \varepsilon_{t2}, \beta_1 > 0, \\ q_t^D = q_t^S, \end{cases} \tag{1}$$

где первое уравнение – функция спроса, второе уравнение – функция предложения, третье уравнение – условие равновесия, p_t - цена товара в момент времени t, ε_{t1} и ε_{t2} - случайные составляющие.

Примером систем одновременных уравнений есть Кейнсианская модель формирования доходов, рассматривается закрытая экономика без государственных расходов:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ y_t = c_t + i_t, \end{cases} \tag{2}$$

где первое уравнение – функция потребления, второе уравнение – макроэкономическое тождество, y_t и i_t - значения совокупного выпуска и инвестиций в момент времени t, ε_t - случайная составляющая.

Одной из возможных нестохастических форм модели равновесия на рынке товаров является такая модель:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t, & (1) \\ \tau_t = \alpha_0 + \alpha_1 y_t, & (2) \\ i_t = \gamma_0 + \gamma_1 r_t, & (3) \\ y_{dt} = y_t + \tau_t, & (4) \\ g_t = \bar{g}, & (5) \\ y_t = c_t + i_t + g_t, & (6) \end{cases} \tag{3}$$

где уравнения (1) − функция потребления; (2) − функция налогов; (3) − функция инвестиций; (4) − функция располагаемого дохода; (5) − государственные расходы; (6) − макроэкономическое тождество; y_t − значения в момент времени t национального дохода; c_t − значения в момент времени t потребления; i_t − значения в момент времени t желаемого объема частных инвестиций; g_t − государственных расходов (в данном случае $g_t = \bar{g} = const$); τ_t − объема налогов; $y_{(t)}$ − располагаемого дохода; r_t − процентной ставки.

Переменные в системах одновременных уравнений делятся на два больших класса: эндогенные переменные, значения которых определяются внутри модели и экзогенные - внешние по отношению к модели, их значения определяются вне модели и поэтому считаются фиксированными. Например, в все переменные в системе (1) являются эндогенными, поскольку они определяются внутри системы.

В кейнсианской модели (2) c_t и y_t оцениваются внутри модели (эндогенные переменные), i_t задается вне модели, следовательно, она является экзогенной переменной. Заметим, что переменные c_t и y_t могут быть выражены через i_t и ε_t, а именно:

$$\begin{cases} y_t = \dfrac{\beta_0}{1-\beta_1} + \dfrac{1}{1-\beta_1} i_t + \dfrac{\varepsilon_t}{1-\beta}, \\ c_t = \dfrac{\beta_0}{1-\beta_1} + \dfrac{\beta_1}{1-\beta_1} i_t + \dfrac{\varepsilon_t}{1-\beta}, \end{cases} \tag{4}$$

где коэффициент $\dfrac{1}{1-\beta_1}$ представляет собой денежный мультипликатор, определяющий на какую величину увеличивается совокупный доход при увеличении объема инвестиций на единицу.

Уравнения, составляющие исходную модель, называются структурными уравнениями модели. В свою очередь их подразделяют на поведенческие уравнения (описывают взаимодействие между переменными) и уравнения-тождества (соотношения, которые должны выполняться во всех случаях, и они не содержат параметров и случайных составляющих).

Уравнения, в которых отражена схема определения эндогенных переменных, называются уравнениями в приведенной форме (приведенными уравнениями). Приведенными уравнениями называются уравнения, в которых эндогенные переменные выражены только через экзогенные или предопределенные переменные, а также случайные составляющие. Примерами таких уравнений являются уравнения системы (4). Предопределенными переменными называются лаговые эндогенные переменные, значения которых определены до рассмотрения соотношения.

11.2. Методы оценивания параметров систем уравнений. Косвенный метод наименьших квадратов (КМНК)

Непосредственное использование МНК для оценки параметров каждого из уравнений системы одновременных уравнений приводит к получению смещенных и несостоятельных оценок. Как правило, это происходит вследствие коррелированности одной или нескольких объясняющих переменных со случайным отклонением. Поэтому применяют другие методы.

Наибольшее распространение получили такие методы оценивания коэффициентов структурной модели: косвенный метод наименьших квадратов; двухшаговый метод наименьших квадратов; трехшаговый метод наименьших квадратов; метод максимального правдоподобия с полной информацией; метод максимального правдоподобия с ограниченной информацией.

Косвенный и двухшаговый методы наименьших квадратов рассматриваются как традиционные методы оценки коэффициентов структурной модели. Эти методы не сложно реализуемы. Косвенный метод наименьших квадратов применяется для идентифицируемой системы одновременных уравнений, а двухшаговый метод наименьших квадратов используется для оценки коэффициентов сверхидентифицируемой модели. Остальные наведенные методы также используются для сверхидентифицируемых систем уравнений.

Метод максимального правдоподобия рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК. Однако при большом числе уравнений системы этот метод приводит к достаточно сложным вычислительным процедурам. Поэтому в качестве модификации служит метод максимального правдоподобия с ограниченной информацией (метод наименьшего дисперсионного отношения). Однако на популярность данного метода в 60-х годах, он был практически вытеснен двухшаговым методом наименьших квадратов (ДМНК), в связи с большой простотой последнего.

Дальнейшим развитием ДМНК является трехшаговый метод наименьших квадратов (ТМНК). Этот метод оценивания пригоден для всх видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным является ДМНК.

Косвенный метод наименьших квадратов (КМНК), основанный на использовании приведенных уравнений. Такое название методу обусловлено вычислением оценок b_1 и b_2 через оценки приведенных уравнений. Оценки b_1 и b_2 полученные по КМНК, являются состоятельными. Следует также отметить, что оценки b_1 и b_2 определяются однозначно, а соответствующее уравнение называется идентифицируемым.

Косвенный метод наименьших квадратов включает в себя такие этапы: 1) исходя из структурных уравнений, строятся уравнения в приведенной форме; 2) оцениваются по МНК параметры уравнений в приведенной форме; 3) на основе найденных оценок, оцениваются параметры структурных уравнений.

Пример. Рассмотри модель «спрос-предложение»:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, \\ q_t^S = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t, \\ q_t^D = q_t^S \end{cases}$$

Здесь первое уравнение – функция спроса, второе уравнение – функция предложения, третье уравнение – условие равновесия, p_t и ϖ_t - цена товара и зарплата в момент времени t соответственно, ε_t и ν_t - случайные составляющие.

Имеются следующие результаты наблюдений.

p	10	15	5	8	4
q	6	6	18	12	8
ϖ	2	6	2	7	4

Найдем оценки параметров этой системы с помощью КМНК.

Обозначим $q_t^D = q_t^S$ через q_t. Тогда

$$\begin{cases} q_t = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, \\ q_t = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t, \end{cases}$$

Приравняем правые части этих уравнений:

$\alpha_0 + \alpha_1 p_t + \varepsilon_{t1} = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t$, отсюда

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2 \varpi_t}{\alpha_1 - \beta_1} + \frac{\nu_t - \varepsilon_t}{\alpha_1 - \beta_1}.$$

Введем обозначения: $\pi_{10} = \dfrac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$, $\pi_{11} = \dfrac{\beta_2}{\alpha_1 - \beta_1}$, $\psi_t = \dfrac{\nu_t - \varepsilon_t}{\alpha_1 - \beta_1}$.

Тогда $p_t = \pi_{10} + \pi_{11} \varpi_t + \psi_t$. Подставляем это выражение для p_t в первое уравнение, раскроем скобки и введем следующее обозначение:

$\pi_{20} = \alpha_0 + \alpha_1 \pi_{10}$, $\pi_{21} = \alpha_1 \pi_{11}$, $\phi_t = \alpha_1 \psi_t + \varepsilon_t$.

Тогда $q_t = \pi_{20} + \pi_{21} \varpi_t + \phi_t$.

Получаем систему из приведенных уравнений:

$$\begin{cases} p_t = \pi_{10} + \pi_{11} \varpi_t + \psi_t \\ q_t = \pi_{20} + \pi_{21} \varpi_t + \phi_t \end{cases}$$

Оценим параметры каждого из этих уравнений при помощи МНК, имеем $\pi_{11} = 0{,}75$; $\pi_{10} = 5{,}25$; $\pi_{21} = -0{,}46$; $\pi_{20} = 11{,}53$.

Тогда $\alpha_1 = \dfrac{\pi_{21}}{\pi_{11}} = -\dfrac{0{,}46}{0{,}75} = -0{,}61$, $\alpha_0 = \pi_{20} - \alpha_1 \pi_{10} = 11{,}53 - (-0{,}61)5{,}25 = 14{,}73$.

Мы не можем оценить коэффициенты β_i на основании полученных результатов. Возникает так называемая проблема идентификации.

11.3. Проблема идентификации

Под проблемой идентификации понимается возможность численной оценки параметров структурных уравнений по оценкам коэффициентов приведенных уравнений.

Исходная система уравнений называется:

а) идентифицируемой, если по оценкам коэффициентов приведенных уравнений можно однозначно определить коэффициенты структурных уравнений (обычно это удается сделать, когда количество уравнений для определения коэффициентов структурных уравнений в точности равно количеству этих коэффициентов);

б) неидентифицируемой (недоопределенной), если по коэффициентам приведенных уравнений можно получить несколько вариантов значений коэффициентов структурных уравнений (обычно это происходит когда количество уравнений для определения коэффициентов структурных уравнений меньше числа определяемых коэффициентов);

в) сверхидентифицируемой (переопределенной), если по оценкам коэффициентов приведенных уравнений невозможно определить коэффициенты структурных уравнений. В данном случае система, связывающая коэффициенты структурных уравнений с коэффициентами приведенных уравнений, является несовместной (обычно в таких случаях число уравнений для оценки коэффициентов структурных уравнений больше числа определяемых коэффициентов).

Для определения идентифицируемости структурных уравнений применяются необходимые и достаточные условия. Пусть система одновременных уравнений включает в себя N уравнений относительно N эндогенных переменных. Система содержит M экзогенных либо предопределнных переменных. Пусть n и m – количество соответственно эндогенных и экзогенных переменных в проверяемом на идентифицируемость уравнении. Переменные, не входящие в данное уравнение, но входящие в другие уравнения системы, называют исключенными переменными (их количество равно $N-n$ для эндогенных и $M-m$ для экзогенных переменных соответственно).

Первое необходимое условие идентифицируемости. Уравнение идентифицируемо, если оно исключает по крайней мере $N-1$ переменную (эндогенную или экзогенную), присутствующую в модели:

$$(N-n)+(M-m)\geq N-1.$$

Второе необходимое условие идентифицируемости. Уравнение идентифицируемо, если количество исключенных из уравнения экзогенных переменных не меньше количества эндогенных переменных в этом уравнении, уменьшенного на единицу:

$$M-m\geq n-1.$$

Знаки равенств в обоих необходимых условиях соответствуют точной идентификации уравнений.

Необходимое и достаточное условие идентифицируемости. В модели, содержащей N уравнений относительно N эндогенных переменных, условие идентифицируемости выполняется тогда и только тогда, когда ранг матрицы, составленной из исключенных их данного уравнений переменных, но входящих в другие уравнения системы, равен $N-1$.

Пример. Исследовать модифицированную модель Кейнса.

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ i_t = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \nu_t, \\ y_t = c_t + i_t + g_t. \end{cases}$$

Здесь g_t - объем государственных расходов. Эндогенные переменные: c_t, i_t, y_t ($N=3$). Экзогенные переменные: g_t, y_{t-1}($M=2$).

Первое уравнение содержит эндогенные переменные c_t, y_t ($n=2$) и не содержит экзогенные переменные ($m=0$).

(N - n) + (M - m)$\geq$ N – 1, то есть (3-2)+(2-0)$\geq$3-1 и 3$\geq$2. Поэтому первое уравнение переопределено.

Второе уравнение содержит эндогенные переменные i_t, y_t ($n=2$) и экзогенную переменную ($m=1$).

(N - n) + (M - m)$\geq$ N – 1, то есть (3-2)+(2-1)$\geq$3-1 и 2=2. Поэтому второе уравнение идентифицируемо.

Третье уравнение содержит эндогенные переменные c_t, i_t, y_t ($n=3$) и экзогенную переменную g_t ($m=1$).

(N - n) + (M - m)$\geq$ N – 1, то есть (3-3)+(2-1)$\geq$3-1 и 1$\geq$2 (ложно). Поэтому третье уравнение сверхидентифицировано.

11.4. Двухшаговый МНК (ДМНК)

Этот метод применяется к переопределенным уравнениям.

Продолжение примера. Первое уравнение исходной системы $c_t = \beta_0 + \beta_1 y_t + \varepsilon_t$ переопределено, то есть по оценкам коэффициентов приведенных уравнений невозможно определить оценки коэффициентов β_0 и β_1. Для переменной y_t строим приведенное уравнение $y_t = \pi_{10} + \pi_{11} y_{t-1} + \pi_{12} g_t + \varpi_t$ (ϖ_t - случайное отклонение), находим с помощью МНК оценки коэффициентов π_{10}; π_{11}; π_{12} и из уравнения получаем оценки $\hat{y}_t$ по экзогенным переменным g_t, y_{t-1}.

Из уравнения $c_t = \beta_0 + \beta_1 \hat{y}_t$ находим оценки коэффициентов β_0 и β_1 с помощью МНК.

Пример. Рассмотрим эконометрическую модель Клейна, составленную для описания экономики США в период с 1921 по 1941 г.г.

Структурная модель содержит три регрессионных уравнения и три тождества:

где: C - фонд потребления (расходы на потребительские товары); I - капиталовложения; W- личные доходы в частном секторе; W'- личные доходы в государственном секторе(зарплата); X - объем производства; T - налоги; K - запасы капитала; P - прибыль; G - государственные расходы на товары и услуги; t - календарное время (год). Все переменные измеряются в миллионах долларов США в ценах 1931 г. Требуется оценить параметры модели b_0 , b_1 , b_2 , b_3 ; b_0 , b'_1 , b'_2 , b'_3 ; b''_0 , b''_1 , b''_2 , b''_3 .

Исходные данные приведены в табл. 11.1. Для переменных P_t , K_t , X_t в скобках приведены данные за 1920 г., так как в модель входят эти же переменные с запаздыванием P_{t-1} , K_{t-1} , X_{t-1} . Значение t_{cp}=1931.

Таблица 11.1

Сводные экономические показатели США за 1921-1941 г.г.

t	*C*	*P*	*W*	*W'*	*I*	*K*	*X*	*G*	*T*
(1920)		*(12.7)*				*(182.8)*	*(44.9)*		
1921	*41.9*	*12.4*	*25.5*	*2.7*	*-0.2*	*182.6*	*45.6*	*3.9*	*7.7*
1922	*45.0*	*16.9*	*29.3*	*2.9*	*1.3*	*184.5*	*50.1*	*3.2*	*3.9*
1923	*49.2*	*18.4*	*34.1*	*2.9*	*5.2*	*189.7*	*57.2*	*2.8*	*4.7*
1924	*50.6*	*19.4*	*33.9*	*3.1*	*3.0*	*192.7*	*57.1*	*3.5*	*3.8*
1925	*52.6*	*20.1*	*35.4*	*3.2*	*5.1*	*197.8*	*63.0*	*3.3*	*5.5*
1926	*55.1*	*19.6*	*37.4*	*3.3*	*5.6*	*203.4*	*64.0*	*3.3*	*7.0*
1927	*56.2*	*19.8*	*37.9*	*3.6*	*4.2*	*207.6*	*64.4*	*4.0*	*6.7*
1928	*57.3*	*21.1*	*39.2*	*3.7*	*3.0*	*210.6*	*64.5*	*4.2*	*4.2*
1929	*57.8*	*21.7*	*41.3*	*4.0*	*5.1*	*215.7*	*67.0*	*4.1*	*4.0*
1930	*55.0*	*15.6*	*37.9*	*4.2*	*1.0*	*216.7*	*61.2*	*5.2*	*7.7*
1931	*50.9*	*11.4*	*34.5*	*4.8*	*-3.4*	*219.3*	*53.4*	*5.9*	*7.5*
1932	*45.6*	*7.0*	*29.0*	*5.3*	*-6.2*	*207.1*	*44.3*	*4.9*	*8.3*
1933	*46.5*	*11.2*	*28.5*	*5.6*	*-5.1*	*202.0*	*45.1*	*3.7*	*5.4*
1934	*48.7*	*12.3*	*30.6*	*6.0*	*-3.0*	*199.0*	*49.7*	*4.0*	*6.8*
1935	*51.3*	*14.0*	*33.2*	*6.1*	*-1.3*	*197.3*	*54.4*	*4.4*	*7.2*
1936	*57.7*	*17.6*	*36.8*	*7.4*	*2.1*	*199.8*	*62.7*	*2.9*	*8.3*
1937	*58.7*	*17.3*	*41.0*	*6.7*	*2.0*	*201.8*	*65.0*	*4.3*	*6.7*
1938	*57.5*	*15.3*	*38.2*	*7.7*	*-1.9*	*199.9*	*60.9*	*5.3*	*7.4*
1939	*61.6*	*19.0*	*41.6*	*7.8*	*1.3*	*201.2*	*60.5*	*6.6*	*8.9*
1940	*65.0*	*21.1*	*45.0*	*8.0*	*3.3*	*204.5*	*75.7*	*7.4*	*9.6*
1941	*69.7*	*23.5*	*53.3*	*8.5*	*4.9*	*209.4*	*88.4*	*13.8*	*11.6*

Сначала были оценены параметры каждого уравнения отдельно стандартным МНК:

$$C_t = 16.24 + 0.193 \cdot P_t + 0.090 \cdot P_{t-1} + 0.796 \cdot (W_t + W'_t) + e_t$$

$$I_t = 10.13 + 0.480 \cdot P_t + 0.333 \cdot P_{t-1} - 0.112 \cdot K_{t-1} + e'_t$$

$$W_t = 1.50 + 0.439 \cdot X_t + 0.146 \cdot X_{t-1} + 0.130 \cdot (t-1931) + e''_t$$

Анализ полученных моделей показал существенные проблемы. Коэффициенты регрессии при переменной P_t и P_{t-1} (прибыль) есть нормы отчислений. Имеем, что коэффициенты перед P_t значительно превышают коэффициенты перед P_{t-1}. Должно же быть наоборот, так как существует некоторая временная задержка между получением прибыли и ее инвестициями. Мы получили смещенные значения параметров - их реальные значения в несколько раз отличаются от вычисленных МНК-оценок. Эконометрическая модель с такими смещенными параметрами - неадекватная, она неверно описывает действительное состояние экономики страны.

Проанализировав систему регрессионных уравнений, обнаруживаем, что в данном примере нарушена первая исходная предпосылка, согласно которой случайные возмущения должны относятся только к результативным признакам и не должны коррелировать с объясняющими переменными. Но в системе одновременных уравнений получается, что совместно зависимые переменные в конечном итоге коррелируют с возмущениями того же уравнения (например, в цепочке уравнений относительно C_t , X_t , P_t следует $e_t \rightarrow C_t \rightarrow X_t \rightarrow P_t$, т.е. корреляция e_t и P_t , чего не должно быть; аналогично, цепочка уравнений относительно I_t , X_t , P_t приводит к непредусмотренной корреляции между e'_t и P_t).

Для оценки параметров системы одновременных уравнений предложена процедура двухшагового регрессионного анализа. Заменим все объясняющие зависимые переменные (в правых частях уравнений системы) на их расчетные значения. Расчетные значения, как известно, не коррелируют со случайными ошибками, поэтому соответствующая предпосылка регрессионного анализа уже не будет нарушена.

В модели имеются переменные: случайные возмущения e_t , e'_t , e''_t ; экзогенные переменные, известные в каждый момент времени t, W'_t , G_t , T_t ; эндогенные переменные, на которые влияют внесистемные переменные C_t , I_t , W_t , X_t , P_t , K_t . Всего таких переменных 6, для их определения имеется 6 уравнений, именно эти показатели мы называем зависимыми объясняющими переменными, когда они входят в уравнение в качестве аргументов. Еще имеются лаговые эндогенные переменные: X_{t-1} , P_{t-1} , K_{t-1} , значения лаговых переменных известны к текущему моменту времени. Внесистемные (экзогенные) и лаговые переменные называются предопределенными, они известны к текущему моменту времени. К ним относятся t, W'_t, G_t, T_t, X_{t-1}, P_{t-1}, K_{t-1} .

Для получения расчетных значений C^R, I^R, W^R составляются линейные модели относительно всех предопределенных переменных; расчетные значения X^R и P^R получаются из соответствующих тождеств:

Следует вычислить расчетные значения и запомнить их в виде новых переменных (предлагается выполнить эти операции на ЭВМ). На втором шаге оцениваются параметры структурной модели:

$$\begin{cases} C_t = 16.55 + 0.017 \cdot P_t^R + 0.216 \cdot P_{t-1} + 0.810 \cdot \left(W_t^R + W_t'\right) + e_t \\ I_t = 20.28 + 0.150 \cdot P_t^R + 0.616 \cdot P_{t-1} - 0.158 \cdot K_{t-1} + e_t' \\ W_t = 1.50 + 0.439 \cdot X_t^R + 0.147 \cdot X_{t-1} + 0.130 \cdot (t - 1931) + e_t'' \end{cases}.$$

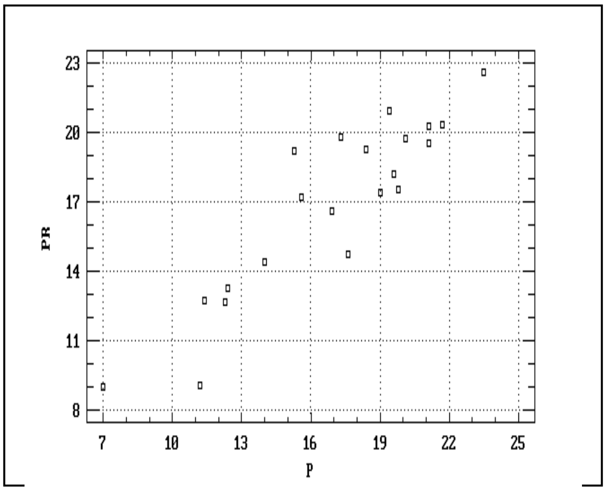


Рис. 11.1. Соответствие между P и P^R

Двухшаговым методом наименьших квадратов получены объяснимые с точки зрения экономики оценки спорных коэффициентов регрессии (сравни их с ранее найденными). Большие изменения в решении получены, в основном, за счет замены P на P^R. На рис. 11.1 показано соответствие между реальными и сглаженными по МНК данными P и P^R.

Коэффициент корреляции между ними равен 0.909. Следовательно, незначительные поправки в исходных данных привели к таким большим смещениям в параметрах модели.

Вопросы для самопроверки:

1) Какие основные причины использования систем одновременных уравнений?
2) Приведите основные типы моделей применения систем одновременных уравнений.
3) Какие виды переменных различают в системах одновременных уравнений? Дать определение их.
4) В чем состоит различие между структурными уравнениями системы и уравнениями в приведенной форме?
5) Почему обычный МНК практически не используется для оценки систем одновременных уравнений?
6) Какие наиболее распространенные методы оценивания коэффициентов структурной модели?
7) В чем состоит суть косвенного метода наименьших квадратов?
8) Назовите причины неидентифицируемости и сверхидентифицируемости систем одновременных уравнений.
9) Приведите необходимые и достаточные условия идентифицируемости систем.
10) В чем состоит суть двухшагового метода наименьших квадратов?

Тема 12. Динамические эконометрические модели

12.1. Общая характеристика динамических эконометрических моделей

12.2. Интерпретация параметров моделей с распределенным лагом

12.3. Изучение структуры лага и выбор вида модели с распределенным лагом

12.1. Общая характеристика динамических эконометрических моделей

В экономическом анализе используются ежегодные, ежеквартальные, ежемесячные, ежедневные данные, представляющие собой значения показателей. Упорядоченные по времени данные называются временными рядами.

Пусть исследуется показатель y, например, прибыль предприятия. Его значения в текущий момент времени t обозначают y_t, а в последующие моменты обозначают $y_{t+1}, y_{t+2}, ..., y_{t+k}, ...$, а в предыдущие моменты обозначаются $y_{t-1}, y_{t-2}, ..., y_{t-k}, ...$. При изучении зависимостей между такими показателями или при анализе их развития во времени в качестве объясняющих переменных используются не только текущие значения переменных, но и некоторые предыдущие по времени значения, а также само время T. Модели такого типа называются динамическими.

Выделяют два типа динамических эконометрических моделей. К моделям первого типа относятся модели авторегрессии и модели с распределенным лагом, в которых значения переменной за прошлые периоды времени (лаговые переменные) непосредственно включены в модель. Модели второго типы учитывают динамическую информацию в неявном виде. В эти модели включены переменные, характеризующие ожидаемый или желаемый уровень результата, или одного из факторов в момент времени t.

Например, на выручку от реализации или прибыль компании текущего периода могут оказывать влияние расходы на рекламу или проведение маркетинговых исследований, сделанных компанией в предшествующие моменты времени. Величину l, характеризующую запаздывание в воздействии фактора на результат, называют лагом, а временные ряды самих факторных переменных, сдвинутые на один или более моментов времени – лаговыми переменными.

Лаговые модели с распределенным лагом имеют вид:

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \varepsilon_t \ .$$

Модели авторегресси имеют вид:

$$y_t = a + b_0 x_t + c_1 y_{t-1} + \varepsilon_t \, .$$

Причин наличия лагов в экономике много, среди них чаще всего выделяют такие: психологические причины, которые обусловлены инерцией в поведении людей; технологические причины, обусловленные инерцией с заменой оборудования, технического перевооружения, инерцией с переходом на новые технологии; институциональные причины, например, контракты между фирмами, трудовые договора требуют определенного постоянства в течении времени; механизмы формирования экономических показателей, например, инфляция является инерционным процессом, денежный мультипликатор (накопление денег в банковской системе) также проявляется на определенном интервале времени.

12.2. Интерпретация параметров моделей с распределенным лагом

Модель с распределенным лагом в общем виде:

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \ldots + b_p x_{t-p} + \varepsilon_t \ .$$

Коэффициент регрессии a характеризует среднее абсолютное изменение в некоторый фиксированный период времени t, без учета воздействия лаговых значений фактора x. Этот коэффициент называют краткосрочным мультипликатором.

Любую сумму коэффициентов $\sum_{j=0}^{h} b_j$, $h < p$ называют промежуточным мультипликатором: в момент $t+1$ совокупное воздействие факторной переменной x_t на результат y_t составит $(b_0 + b_1)$; в момент времени $t+2$ соответственно $(b_0 + b_1 + b_2)$ и т.д. Сумму всех b_j называют долгосрочным мультипликатором $b_0 + b_1 + b_2 + ... + b_l = b$ и он характеризует изменение y под воздействием единичного изменения x в каждом из рассматриваемых периодов времени. Рассчитываются относительные коэффициенты модели с распределенным лагом: $\beta_j = \frac{b_j}{b}$. Если все коэффициенты b_j имеют одинаковые знаки, то для любого i-го значения $0 < \beta_j < 1$; $\sum_{j=0}^{l} \beta_j = 1$. Относительные коэффициенты β_j являются весами для соответствующих коэффициентов b_j, каждый из них измеряет долю общего изменения результативного признака в момент времени $t + j$. Зная величину β_j можно определить средний лаг и медианный лаг.

Средний лаг: $\bar{l} = \sum_{j=0}^{l} j \cdot \beta_j$ представляет собой средний период, в течение которого будет происходить изменение результата под воздействием изменения фактора в момент времени t. Небольшая величина среднего лага свидетельствует об относительно быстром реагировании результатов на изменение фактора, а большая величина – о том, что воздействие фактора на результат состоится через длительный период времени. Медианный лаг – это величина лага, для которого $\sum_{j=0}^{l_{Me}} \beta_j \approx 0{,}5$. Это период, в течение которого с момента времени t будет реализована половина общего воздействия фактора на результат.

В модели авторегрессии $y_t = a + b_0 x_t + c_1 y_{t-1} + \varepsilon_t$, как и в модели с распределенным лагом, b_0 характеризует краткосрочное изменение y_t под воздействием изменения x_t на 1 ед. К моменту времени $t+1$ результат y_t изменится под воздействием изменения фактора в момент времени t на b_0, а y_{t-1} под воздействием своего изменения в непосредственно предшествующий момент времени на c_1. Абсолютное изменение результата в момент времени $t+1$ составит $(b_0 c_1)$, а в момент времени $t+2$ абсолютное изменение составит $(b_0 c_1^2)$ и т.д. Долгосрочный мультипликатор в модели авторегрессии рассчитывается как сумма краткосрочного и промежуточного мультипликатора:

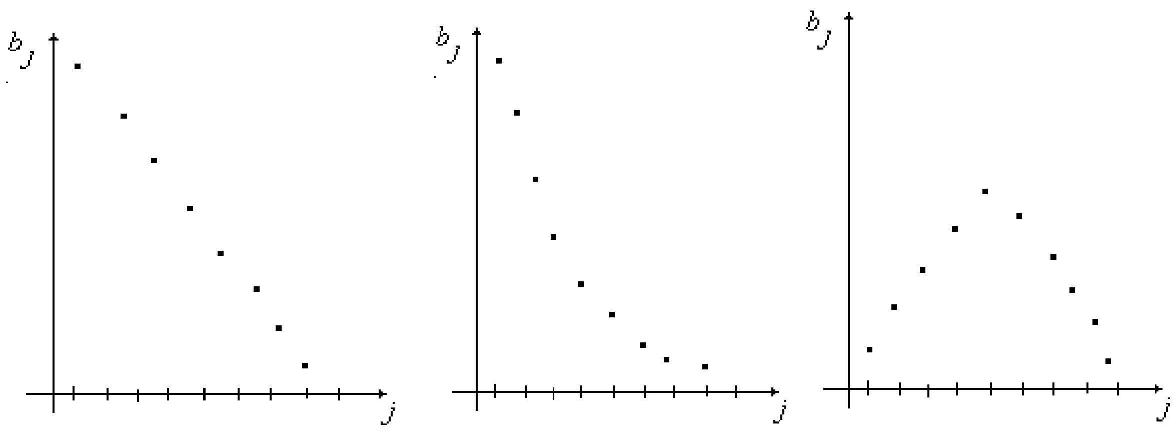
$$b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 + \ldots .$$

Учитывая $|c_1| < 1$, имеем $b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 + \ldots = \dfrac{b_0}{1 - c_1}$.

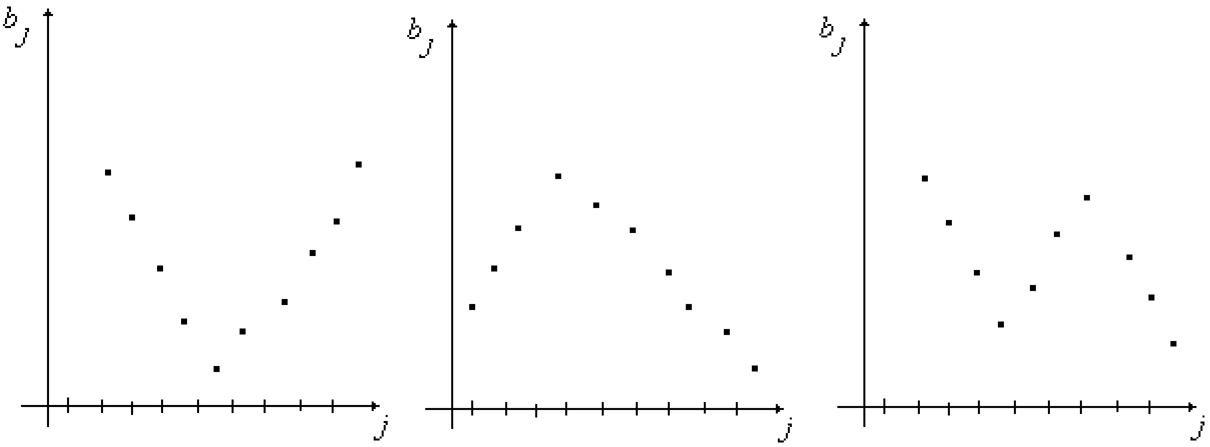
Такая интерпретация коэффициентов модели авторегрессии и расчет долгосрочного мультипликатора основаны на предпосылке о наличии бесконечного лага в воздействии текущего значения зависимой переменной на ее будущие значения.

12.3. Изучение структуры лага и выбор вида модели с распределенным лагом

Если построить график зависимости коэффициентов регрессии от величины лага, можно получить графическое изображение структуры лага, или распределение во времени воздействия факторной переменной на результат. Структура лага может быть различной (рис. 12.1). Аналогично графический анализ структуры лага можно проводить и с помощью относительных коэффициентов регрессии β_j. Основная трудность в выявлении структуры лага состоит в том, как получить значения параметров b_j или β_j. Обычный МНК редко бывает полезным в этих целях. Поэтому в большинстве случаев предположение о структуре лага основаны на общих положениях экономической теории.



а) б) в)



г) д) е)

Рис. 12.1. Графическое изображение структуры лага: а) линейная, б) геометрическая, в) перевернутая V-образная, г), д), е) полиномиальная.

Вопросы для самопроверки:

1) Какие модели называются динамическими?
2) Какие типы динамических моделей различают?
3) Что называется лагом? Какие переменные называются лаговыми?
4) Какие причины в экономике существования лагов?
5) Что называется краткосрочным мультипликатором в моделях с распределенным лагом?
6) Что называется промежуточным и долгосрочным мультипликаторами в моделях с распределенным лагом?
7) Что характеризует средний лаг в моделях с распределенным лагом?
8) Что характеризует медианный лаг в моделях с распределенным лагом?
9) Что называется краткосрочным мультипликатором в модели авторегрессии?
10) Что характеризует средний лаг в модели авторегрессии?
11) Что характеризует медианный лаг в модели авторегрессии?
12) Что характеризует средний лаг в модели авторегрессии?
13) Как изучить структуру лага в моделях с распределенным лагом?

Тема 13. Методы разработки динамических эконометрических моделей

13.1. Метод Алмон
13.2. Метод Койка

13.1. Метод Алмон

Метод Алмон основан на теореме Вейэрштрасса, согласно которой предполагается, что весовые коэффициенты подчиняются полиномиальному распределению. На рис.13.1 представлены распределения весовых коэффициентов.

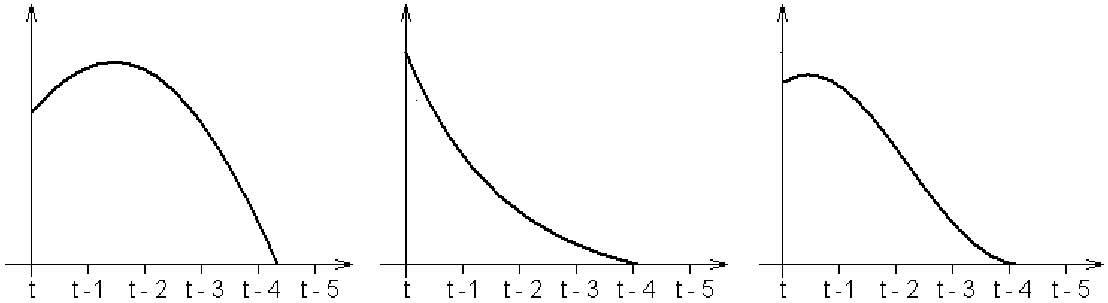


Рис.13.1. Априорные представления о распределении весовых коэффициентов

Предполагается, что подобные зависимости можно аппроксимировать полиномом небольшой степени второй или третьей. Выбор функции остается за исследователем и может быть разработан на основе экспериментов. Далее, исследователь должен выбрать число лаговых значений объясняющей переменной. И это число, определенное в результате эксперимента, направлено на получение хорошего описания имеющихся данных. Другими словами, вместо использования обычного МНК оценивания параметров модели

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{i-1} + b_2 \cdot x_{t-2} + \ldots + b_k \cdot x_{t-k},$$

предлагается изначально аппроксимировать поведение коэффициентов регрессии полиномом небольшой степени:

$$b_i = a_0 + a_1 \cdot i + a_2 \cdot i^2 + \ldots a_m \cdot i^m, \quad (m \le k)$$

и временно перейти к новым переменным

$$\begin{aligned}
Z_0 &= x_t + x_{t-1} + x_{t-2} + \ldots + x_{t-k}, \\
Z_1 &= \quad x_{t-1} + 2 \cdot x_{t-2} + \ldots + k \cdot x_{t-k}, \\
Z_2 &= \quad x_{t-1} + 2^2 \cdot x_{t-2} + \ldots + k^2 \cdot x_{t-k}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
Z_m &= \quad x_{t-1} + 2^m \cdot x_{t-2} + \ldots + k^m \cdot x_{t-k}.
\end{aligned}$$

Далее необходимо вычислить параметры модели относительно новых переменных

$$y_t = a + a_0 \cdot Z_0 + a_1 \cdot Z_1 + a_2 \cdot Z_2 + \ldots + a_m \cdot Z_m$$

и проделать обратный переход к первоначальной модели

$$\begin{aligned}
b_0 &= a_0, \\
b_1 &= a_0 + a_1 + a_2 + \ldots + a_m, \\
b_2 &= a_0 + 2 \cdot a_1 + 2^2 \cdot a_2 + \ldots + 2^m \cdot a_m, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
b_k &= a_0 + k \cdot a_1 + k^2 \cdot a_2 + \ldots + k^m \cdot a_m.
\end{aligned}$$

Если $m = k$ (степень аппроксимирующего полинома равняется максимальному лагу), то в конце будут получены МНК-оценки параметров изначальной модели, которые можно было получить проще – непосредственным использованием МНК.

Если $m < k$, то можно получить непредвиденные результаты. Для того, чтобы предотвратить нереальных выводов, рекомендуется рассмотреть графики зависимостей $b_k = f(k)$ и задать самые подходящие значения m и k. Однако тогда теряется объективность метода. Традиционно находят преимущества метода Алмон в то время, когда имеем существенные недочеты. Объем вычислительной работы значительно больший, новые переменные связаны тесными мультиколлинеарными связями, допускаются не контролированные ошибки неудачной аппроксимации, теряется научная объективность, отсутствуют оценки значимости b_i .

Пример. В табл.1 приведено за 24 периода объемы реализации (У) и цены за единицу (Х) продукции завода «Укрэлектромаш» (г. Харьков). Составим лаговую модель

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{i-1} + b_2 \cdot x_{t-2} + \ldots + b_k \cdot x_{t-k}$$

где максимальный лаг k не может быть большим 25% от числа периодов, то есть $k \le 6$.

Для иллюстрации, обычным МНК были вычислены параметры моделей для $k = 4, 5, 6$ вместе со всеми их статистическими характеристиками. На рис. 13.2 приведены графики поведения параметров модели b_i в зависимости от принятого максимального лага k. Видим, что представленные графики отличаются от графиков рис. 13.1.

Методом Алмон были вычислены все комбинации моделей для $k = 4, 5, 6$ і $2 \le m \le k$.

У всех комбинациях, де $m = k$ были получены обычные МНК-оценки параметров, но без статистических оценок их значимости. Для примера рассмотри результаты вычислений при $k = 4$ і $m = 2, 3$ (кстати, это самый распространенный вариант вычислений по методу Алмон).

В табл. 13. 1 приведены значения лаговых переменных, на основании каких вычислены значения новых переменных Z_0, Z_1, Z_2, Z_3. Потеряно $k = 4$ наблюдений (17,7%). Значения Z_j очень быстро возрастают с номером j. Между этими переменными существует очень сильная корреляция, например, $r_{z_2 z_3} = 0{,}999$.

Таблица 13.1

Объемы реализации (У) и цены за единицу (Х) продукции «Укрэлектромаш» (г. Харьков)

Период	$У_t$	X_t	X_{t-1}	X_{t-2}	X_{t-3}	X_{t-4}	Z_0	Z_1	Z_2	Z_3
1	2141	187,74								
2	2934	180,63	187,74							
3	3371	173,79	180,63	187,74						
4	5448	158,34	173,79	180,63	187,74					
5	7200	150,66	158,34	173,79	180,63	187,74	851,16	1798,77	5483,01	18441,03
6	5497	157,82	150,66	158,34	173,79	180,63	1008,98	2649,93	9931,71	41137,53
7	5410	159,35	157,82	150,66	158,34	173,79	1168,33	3658,91	16240,55	79891,43
8	4089	171,39	159,35	157,82	150,66	158,34	1151,98	3513,06	15527,44	76363,32
9	1492	200,87	171,39	159,35	157,82	150,66	1172,22	3400,63	14854,67	72680,71

10	934	199,53	200,87	171,39	159,35	157,82	1197,96	3356,32	14312,44	69008,86
11	907	198,51	199,53	200,87	171,39	159,35	1238,13	3445,9	14464,38	68902,48
12	2630	183,20	198,51	199,53	200,87	171,39	1270,67	3629,41	15211,97	72195,07
13	3033	179,06	183,20	198,51	199,53	200,87	1291,91	3795,34	16008,28	75857,62
14	4192	170,12	179,06	183,20	198,51	199,53	1302,68	3971,8	17082,72	81903,34
15	4226	164,34	170,12	179,06	183,20	198,51	1295,63	4074,75	17930,89	87582,81
16	5047	160,63	164,34	170,12	179,06	183,20	1255,39	3964,29	17533,39	85996,95
17	1067	191,39	160,63	164,34	170,12	179,06	1247,25	3822,97	16940,39	83306,59
18	1008	196,10	191,39	160,63	164,34	170,12	1244,84	3680,65	16106,59	78754,99
19	9572	151,75	196,10	191,39	160,63	164,34	1213,39	3643,09	15735,93	76523,95
20	4259	167,87	151,75	196,10	191,39	160,63	1202,2	3603,06	15461,56	74456,82
21	1490	201,98	167,87	151,75	196,10	191,39	1234,06	3614,42	15534	74501,72
22	446	200,32	201,98	167,87	151,75	196,10	1270,04	3698,1	15944,24	76812,42
23	560	208,51	200,32	201,98	167,87	151,75	1317,92	3843,73	16739,61	81913,39
24	1256	205,55	208,51	200,32	201,98	167,87	1332,08	3821,92	16366,88	79334,56

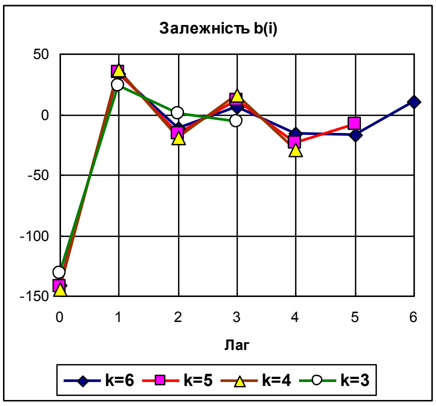


Рис. 13.2. Зависимость параметров модели от лага

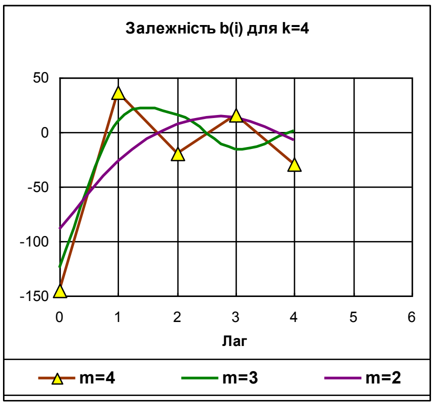


Рис. 13.3. Качество аппроксимации параметров модели

На рис. 13.3 показано качество аппроксимации зависимости b_i полиномами 2-й и 3-й ступеней (полином 4-й ступени точно воспроизводит 5 заданных точек). Как видим, аппроксимация не очень удачная. Для сравнения приведем уравнения регрессий для k = 4 и m = 4, 3, 2.

$m = 4$: $y_t = 28537 - 144{,}7x_t + 36{,}4x_{t-1} - 19{,}3x_{t-2} + 16{,}3x_{t-3} - 29{,}1x_{t-4}$

$m = 3$: $y_t = 23887 - 124{,}9x_t + 9{,}5x_{t-1} + 16{,}3x_{t-2} - 16{,}1x_{t-3} + 0{,}9x_{t-4}$

$m = 2$: $y_t = 21863 - 89{,}6x_t - 27{,}1x_{t-1} + 7{,}4x_{t-2} + 13{,}8x_{t-3} - 7{,}9x_{t-4}$

В этих моделях совпадают даже знаки параметров.

Это еще раз подтверждает преувеличенное значение метода Алмон. С вычислительной точки зрения никаких преимуществ не имеем, поскольку полиномы Алмон также тесно коррелируют между собой, как и изначальные лаговые переменные.

13.2. Метод Койка

Метод Алмон используется для построения модели с конечной величиной лага l . Для оценки параметров модели с бесконечною величиной лага рекомендуется метод Койка.

Л.М. Койк предположил, что существует некоторый постоянный темп λ ($0 < \lambda < 1$) уменьшения во времени лаговых воздействий фактора на результат:

$$b_j = b_0 \cdot \lambda; \quad j = 0,1,2,...$$

Чем ближе λ к 0, тем выше темп снижения воздействия фактора на результат во времени и тем больше доля воздействия на результат приходится на текущие значения фактора x_i .

Выразим с помощью $b_j = b_0 \cdot \lambda; \quad j = 0,1,2,...$ все коэффициенты b_j в модели $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$ через b_0, λ:

$$y_t = a + b_0 x_t + b_0 \lambda x_{t-1} + b_0 \lambda^2 x_{t-2} + ... + \varepsilon_t .$$

Тогда для периода $(t-1)$ модель можно записать следующим образом:

$$y_{t-1} = a + b_0 x_{t-1} + b_0 \lambda x_{t-2} + b_0 \lambda^2 x_{t-3} + ... + \varepsilon_{t-1} .$$

Умножим обе части модели на λ:

$$\lambda \cdot y_{t-1} = \lambda \cdot a + b_0 \lambda \cdot x_{t-1} + b_0 \lambda^2 x_{t-2} + b_0 \lambda^3 x_{t-3} + ... + \lambda \cdot \varepsilon_{t-1} .$$

Вычтем данное выражение из $y_t = a + b_0 x_t + b_0 \lambda x_{t-1} + b_0 \lambda^2 x_{t-2} + ... + \varepsilon_t$:

$y_t - \lambda \cdot y_{t-1} = a - \lambda \cdot a + b_0 x_t + \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$, после преобразования получаем модель Койка $y_t = a(1-\lambda) + b_0 x_t + \lambda \cdot y_{t-1} + u_t$, где $u_t = \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$.

Полученная модель есть модель двух факторной линейной регрессии, точнее авторегрессии. Определив ее параметры, найдем λ и оценки параметров a и b_0 исходной модели. Далее с помощью соотношений $b_j = b_0 \cdot \lambda; \quad j = 0, 1, 2,...$ определяют параметры $b_1, b_2, ...$ модели $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$.

Отметим, что применение обычного МНК к оценке параметров модели $y_t = a(1-\lambda) + b_0 x_t + \lambda \cdot y_{t-1} + u_t$ приведет к получению смещенных оценок ее параметров ввиду наличия в этой модели в качестве фактора лаговой результативной переменной y_{t-1}.

Описанный алгоритм получил название преобразование Койка. Это преобразование позволяет перейти от модели с бесконечными распределенными лагами к модели авторегрессии, содержащей две независимые переменные x_t и y_{t-1}.

Несмотря на бесконечное число лаговых переменных в модели $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$, геометрическая структура лага позволяет определить величины среднего и медианного лагов в модели Койка. Поскольку сумма коэффициентов регрессии в данной модели есть сумма геометрической прогрессии, т.е.

$$\sum_{j=0}^{\infty} b_j = b_0 + b_0\lambda + b_0\lambda^2 + b_0\lambda^3 + \ldots = b_0(\lambda + \lambda^2 + \lambda^3 + \ldots) = \frac{b_0}{1-\lambda},$$

то средний лаг определяется как

$$\bar{l} = \frac{\sum_{j=0}^{\infty} j \cdot b_j}{\sum_{j=0}^{\infty} b_j} = \frac{b_0\lambda \cdot (\lambda + \lambda^2 + \lambda^3 + \ldots)}{\frac{b_0}{1-\lambda}} = \frac{b_0 \cdot \lambda \frac{1}{(1-\lambda)^2}}{\frac{b_0}{1-\lambda}} = \frac{\lambda}{1-\lambda}.$$

Заметим, что при $\lambda = 0{,}5$ средний лаг $T = 1$, а при $l < 0{,}5$ средний лаг $T < 1$, т.е. воздействие фактора на результат в среднем занимает менее одного периода времени. Величину $(1-\lambda)$ интерпретируют обычно как скорость, с которой происходит адаптация результата во времени к изменению факторного признака. Для расчета медианного лага необходимо выполнение следующего условия:

$$\sum_{j=0}^{l_{Me}-1} b_j = \sum_{j=0}^{l_{Me}-1} \frac{b_j}{\sum_{j=0}^{\infty} b_j} = \sum_{j=0}^{l_{Me}-1} \frac{b_0 \cdot \lambda^j}{\frac{b_0}{1-\lambda}} = \sum_{j=0}^{l_{Me}-1} \lambda^j (1-\lambda) = 0{,}5.$$

Поэтому медианный лаг в модели Койка равен: $l_{Me} = \dfrac{\ln 0{,}5}{\ln \lambda}$.

Пример. Использовав данные экономики Канады в течении 1961 – 1981 г.г. , ученые Джеффри Сакс и Майкл Бруно построили модель:

$$U_t = \delta_0 + \delta_1 \cdot U_{t-1} + \delta_2 \cdot t + \delta_3 \cdot w_t + \varepsilon_t,$$

где U_t, U_{t-1} – уровень безработицы в периоды t и $t-1$;

w_t – превышение реальной заработной платы сравнительно с ее уровнем в условиях полной занятости (значения получены с помощью вычислений);

t – время;

$\delta_0, \delta_1, \delta_2, \delta_3$ – параметры модели;

ε_t – ошибка;

$$U_t = \delta_0 + 0{,}63 \cdot U_{t-1} + 0{,}07 \cdot t + 15{,}72 \cdot w_t, \quad R^2 = 0{,}85.$$
$$t = 5{,}46 \qquad t = 2{,}82 \quad t = 2{,}23$$

Переменная w_t в модели обозначает спрос на работу. Если предусмотреть, что переменная w_t осуществляет влияние на уровень безработицы с бесконечным лагом в условиях геометрической структуры лага, то соответственно к методу Койка получим такую модель с распределенным лагом:

$$U_t = a + b_0 w_t + b_0 \cdot \lambda \cdot w_{t-1} + b_0 \cdot \lambda^2 \cdot w_{t-2} + \ldots + c \cdot t + \varepsilon_t.$$

Эта модель отличается от $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + \varepsilon_t$ тем, что, кроме текущего и лагового значений факторного признака, она учитывает фактор времени t. Соответствующие алгебраические преобразования согласно метода Койка приводят к модели авторегрессии вида:

$$U_t = \left(a \cdot (1-\lambda) + \lambda \cdot c\right) + (1-\lambda) U_{t-1} + c \cdot (1-\lambda) t + b_0 w_t + u_t,$$

то есть $\delta_0 = a \cdot (1-\lambda) + \lambda \cdot c;\quad \delta_1 = 1-\lambda;\quad \delta_2 = c \cdot (1-\lambda);\ \delta_3 = b_0$.

В модели Джеффри Сакса и Майкла Бруно $\lambda = 0{,}63$. При этом параметры модели Койка:

$$c = \frac{0{,}07}{1-0{,}63} = 0{,}189;$$

$$a = \frac{\delta_0}{1-0{,}63} + 0{,}189 \cdot 0{,}63 = 0{,}119 + 2{,}703 \cdot \delta_0;$$

$$b_0 = 15{,}72;$$

$$b_1 = 15{,}72 \cdot 0{,}63 = 9{,}904;$$

$$b_2 = 15{,}72 \cdot (0{,}63)^2 = 6{,}239;$$

$$b_3 = 15{,}72 \cdot (0{,}63)^3 = 3{,}931 \text{ и т.д.}$$

Модель Койка приобретает такой вид:

$$U_t = (0{,}119 + 2{,}703 \cdot \delta_0) + 15{,}72 \cdot w_t + 9{,}904 \cdot w_{t-1} + 6{,}239 \cdot w_{t-2} + 3{,}931 \cdot w_{t-3} + \\ + \ldots + 0{,}189t.$$

Средний лаг равен $T = \frac{0{,}63}{1-0{,}63} = 1{,}703$; величина медианного лага: $l_{Me} = \frac{\ln 0{,}5}{\ln 0{,}63} = \frac{-0{,}69314}{-0{,}46203} = 1{,}500203$. Отсюда можно сделать выводы, что в среднем влияние отклонений между реальной заработной платой в экономике Канады и ее величиной в условиях полной занятости проявляется в течении относительно краткого периода времени, это 1,7 года, при этом половина влияния реализуется в течении первых 1,5 годов с моменты изменения w_t.

Пример. Вычислить модель с распределенными лагами для описания затрат на жилье (y) в зависимости от уровня доходов (x) и относительных цен текущего (p) и нескольких предыдущих периодов (табл. 13.2); преобразовать эту модель в авторегрессионную методом Койка; оценить ее параметры и сделать выводы относительно кратко- и долгосрочного влияния объясняющих переменных.

Для демонстрации были вычислены модели, где $x_{t-1}, p_{t-1}, x_{t-2}, p_{t-2}$ лаговые переменные, учитывающие сдвиг в один и два периоды:

1) $\ln y_p = -1{,}51 + 1{,}81 \ln x - 0{,}35 \ln p;\ R^2 = 0{,}991;$

2) $ln\, y_p = 0{,}86 + 1{,}09\, ln\, x_{t-1} - 0{,}73\, ln\, p_{t-1};\ R^2 = 0{,}993;$

$$t_{b_0} = 0{,}5;\quad t_{b_{xt-1}} = 19{,}4;\quad t_{b_{pt-1}} = 2{,}2;$$

Таблица 13.2

Значения переменных

Рік	y_t	x_t	p_t	$\ln y_t$	$\ln y_{t-1}$	$\ln x_t$	$\ln p_{t-1}$	$\ln x_{t-1}$	$\ln p_{t-1}$	$\ln x_{t-2}$	$\ln p_{t-2}$
1960	64,0	489,7	104,5	4,1589	4,1092	6,1938	4,6492	6,1732	4,6492	6,1327	4,6501
1961	67,0	503,8	105,1	4,2047	4,1589	6,2222	4,6549	6,1938	4,6492	6,1732	4,6492
1962	70,7	524,9	105,0	4,2584	4,2047	6,2632	4,6540	6,2222	4,6549	6,1938	4,6492
1963	74,0	542,3	104,8	4,3041	4,2584	6,2958	4,6521	6,2632	4,6540	6,2222	4,6549
1964	77,4	580,8	104,5	4,3490	4,3041	6,3644	4,6492	6,2958	4,6521	6,2632	4,6540
1965	81,6	616,3	104,0	4,4018	4,3490	6,4237	4,6444	6,3644	4,6492	6,2958	4,6521
1966	85,3	646,8	102,6	4,4462	4,4018	6,4720	4,6308	6,4237	4,6444	6,3644	4,6492
1967	89,1	673,5	102,2	4,4898	4,4462	6,5125	4,6269	6,4720	4,6308	6,4237	4,6444
1968	93,5	701,3	100,9	4,5380	4,4898	6,5529	4,6141	6,5125	4,6269	6,4720	4,6308
1969	98,4	722,5	100,0	4,5890	4,5380	6,5827	4,6052	6,5529	4,6141	6,5125	4,6269
1970	102,0	751,6	99,6	4,6250	4,5890	6,6222	4,6012	6,5827	4,6052	6,5529	4,6141
1971	106,4	779,2	100,0	4,6672	4,6250	6,6583	4,6052	6,6222	4,6012	6,5827	4,6052
1972	112,5	810,3	100,0	4,7230	4,6672	6,6974	4,6052	6,6583	4,6052	6,6222	4,6012
1973	118,2	865,3	99,1	4,7724	4,7230	6,7631	4,5961	6,6974	4,6052	6,6583	4,6052
1974	124,2	858,4	95,1	4,8219	4,7724	6,7551	4,5549	6,7631	4,5961	6,6974	4,6052
1975	128,3	875,8	93,3	4,8544	4,8219	6,7751	4,5358	6,7551	4,5549	6,7631	4,5961
1976	134,9	906,8	93,7	4,9045	4,8544	6,8099	4,5401	6,7751	4,5358	6,7551	4,5549
1977	141,3	942,9	94,5	4,9509	4,9045	6,8490	4,5486	6,8099	4,5401	6,7751	4,5358
1978	148,5	988,8	94,7	5,0006	4,9509	6,8965	4,5507	6,8490	4,5486	6,8099	4,5401
1979	154,8	1015,5	93,8	5,0421	5,0006	6,9231	4,5412	6,8965	4,5507	6,8490	4,5486
1980	159,8	1021,6	93,0	5,0739	5,0421	6,9291	4,5326	6,9231	4,5412	6,8965	4,5507
1981	164,8	1049,3	94,2	5,1047	5,0739	6,9559	4,5454	6,9291	4,5326	6,9231	4,5412
1982	167,5	1058,3	96,7	5,1210	5,1047	6,9644	4,5716	6,9559	4,5454	6,9291	4,5326
1983	171,3	1095,4	99,2	5,1434	5,1210	6,9989	4,5971	6,9644	4,5716	6,9559	4,5454

3) $ln\, y_p = 2{,}31 + 1{,}02\, ln\, x_{t-2} - 0{,}95\, ln\, p_{t-2}$; $R^2 = 0{,}996$;

$$t_{b_0} = 1{,}4; \quad t_{b_{xt-2}} = 20{,}6; \quad t_{b_{pt-2}} = 3{,}3.$$

Таким образом, вычисленные модели почти не имеют отличий: значения эластичностей затрат по доходам почти одинаковые; все уравнения значимые в целом и имеют почти одинаковые коэффициенты детерминации. Поэтому не имеет ни каких предпочтения та или иная модели. Это не случайно, поскольку лаговые переменные обычно тесно коррелируют. Поэтому коэффициенты регрессии в любом из этих уравнений фактически отображают суммарные эффекты по нескольким временным периодам. Правильная модель должна учитывать эффекты каждого периода отдельно, то есть необходимо разработать модель с распределенным лагом:

4) $ln\, y_p = -1{,}51 + 1{,}81 ln\, x_t - 0{,}35 ln\, p_t$;

(t_b) (0,8) (21,0) (1,1)

$$t_{b_0} = 0{,}8; \quad t_{b_x} = 21{,}0; \quad t_{b_p} = 1{,}1$$

5) $ln\, y_p = 0{,}17 + 0{,}22\, ln\, x_t + 1.00\, ln\, p_t + 0{,}88\, ln\, x_{t-1} - 1{,}60\, ln\, p_{t-1}$;

$$t_{b_0} = 0{,}1; \quad t_{b_{ln\,x}} = 0{,}8; \quad t_{b_{ln\,p}} = 2{,}8; \quad t_{b_{ln\,xt-1}} = 2{,}9; \quad t_{b_{ln\,pt-1}} = 4{,}0;$$

6)

$$\ln y_p = 2{,}14 + 0{,}31 \ln x_t - 0{,}16 \ln p_t + 0{,}18 \ln x_{t-1} - 0{,}65 \ln p_{t-1} +$$

$$+ 0{,}55 \ln x_{t-2} - 1{,}44 \ln p_{t-2};$$

$$t_{b_0} = 1{,}2; \quad t_{b_{ln\,x}} = 1{,}2; \quad t_{b_{ln\,p}} = 0{,}3; \quad t_{b_{ln\,xt-1}} = 0{,}4; \quad t_{b_{ln\,pt-1}} = 0{,}7;$$

$$t_{b_{ln\,xt-2}} = 1{,}6; \quad t_{b_{ln\,pt-2}} = 2{,}5.$$

Все три модели значимые в целом и имеют почти одинаковые коэффициенты детерминации $R^2 = 0{,}991 \div 0{,}997$. Однако с увеличением членов коэффициенты регрессии стают все менее значимые и все более не устойчивыми. Это объясняется мультиколлинеарностью – тесными корреляционными связями между лаговыми переменными.

С помощью преобразований Л. Койка получим модель:

$$ln\, y_p = 0{,}50 + 0{,}15\, ln\, x_t - 0{,}16\, ln\, p_t + 0{,}845\, ln\, y_{t-1}; \; R^2 = 0{,}9996$$

$$t_{b_0} = 1{,}3; \quad t_{b_{ln\,x}} = 3{,}1; \quad t_{b_{ln\,p}} = 2{,}4; \quad t_{b_{ln\,yt-1}} = 22{,}4.$$

С помощью данной модели можно оценить кратко- и долгосрочные эффекты. В краткосрочном аспекте (для текущего периода) значения y_{t-1} необходимо рассматривать как фиксированное, тогда эластичность затрат по доходам и цены будут равняться коэффициентам регрессии: $b = 0{,}15; \quad c = -0{,}16$. В долгосрочном аспекте, когда x, p, y приравниваются к свои равновесовым значениям (тогда $y_{t-1} = y_t$), оказывается, что долгосрочное влияние x на y равняется величине $\frac{b}{1-q}=\frac{0{,}15}{1-0{,}845}=0{,}96$, а долгосрочное влияние p на y равняется $\frac{c}{1-q}=\frac{-0{,}16}{1-0{,}845}=-1{,}02$. Эти числа близки к значимым коэффициентам регрессии модели (1) – (3).

Преобразование Койка основывается на нескольких отягощающих предположениях, что коэффициенты регрессии при лаговых объясняющих переменных экспоненциально убывают, начиная с первого члена. Иногда следует предусмотреть, что изменение зависимой переменной у ответ на изменение объясняющей переменной сначала небольшая, а потом возрастает со временем и только потом начинает уменьшаться. Чтобы учесть этот возможный эффект, примем гипотезу об экспоненциальном убывании весовых коэффициентов не с первого, а со второго члена. Тогда преобразование Койка приведет к модели:

$$ln\,y = a + (b\,ln\,x + c\,ln\,p) + (b_0\,ln\,x_t + c_0\,ln\,p_t) + qy_t + u,$$

где $b_0 = b_1 - bq;\; c_0 = c_1 - cq$.

Весовые коэффициенты для цен прошлых лет вычисляются аналогично. Параметры b_0, c_0 уточняют поведение первого члена (b_1, c_1).

Вопросы для самопроверки:

1) В чем суть метода Алмон?

2) При какой структуре лага применим метод Алмон?

3) В чем отличие применения метода Алмон при построении реальной модели с распределенным лагом?

4) В чем суть преобразования Койка?

5) При какой структуре лага применим метод Койка?

6) В чем отличие применения метода Койка при построении реальной модели с распределенным лагом?

Тема 14. Моделирование одномерных временных рядов

14.1. Основные элементы временного ряда

14.2. Автокорреляция уровней временного ряда и выявление его структуры

14.3. Сглаживание временных рядов с помощью скользящих средних

14.1. Основные элементы временного ряда

Модели, построенные на основе данных, характеризующих один объект за ряд последовательных моментов (периодов) времени, называются моделями временных рядов.

Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- факторы, формирующие тенденцию ряда;
- факторы, формирующие циклические колебания ряда;
- случайные факторы.

Каждый уровень временного ряда формируется под воздействием тенденции, сезонных колебаний и случайной компоненты. Процедура идентификации называется декомпозицией. Каждая компонента идентифицируется отдельно. Затем вклады каждой компоненты комбинируются с целью получения прогнозов будущих значений временных рядов. Методы декомпозиции используются как для кратковременных, так и для долговременных прогнозов. С их помощью можно отображать рост или спад, лежащий в основе ряда, или корректировать значения ряда, исключая из них одну или несколько компонент.

Следует отметить, что в последнее время к прогнозам, сделанным на основе метода декомпозиции, относятся как к промежуточным, а сам метод рассматривают как инструмент достижения понимания временных рядов.

Метод декомпозиции предполагает выделения компонент: трендовой, циклической, сезонной и случайной. Тренд – это компонента, представляющая основной рост (спад) во временном ряду. Трендовая компонента образовывается под влиянием постоянного изменения фактора и обозначается буквой T. Циклическая компонента – последовательность волнообразных флуктуаций или циклы более одного года. Изменение экономических условий обычно происходит циклически. Обозначается циклическая компонента буквой C. На практике идентифицировать цикл сложно, он часто кажется частью тренда. В этом случае компоненту называют трендово-циклической и обозначают буквой T. Сезонные изменения обычно присутствуют в квартальных, месячных и недельных данных. Под сезонными вариациями понимаются изменения с более или менее стабильной структурой, имеющие годовую цикличность. Сезонная компонента обозначается буквой S.

Случайная компонента обусловлена действием множеством разнообразных событий, которые сами по себе несущественны, но совместно могут дать значительный эффект E. При различных сочетаниях в изучаемом явлении или процессе эти компоненты могут принимать различные формы. На рис. 14.1 а) представлен гипотетический временной ряд, содержащий возрастающую тенденцию; на рис. 14.1 б) – гипотетический временной ряд содержащий только сезонную компоненту; на рис. 1 в) – гипотетический временной ряд содержащий только сезонную компоненту.

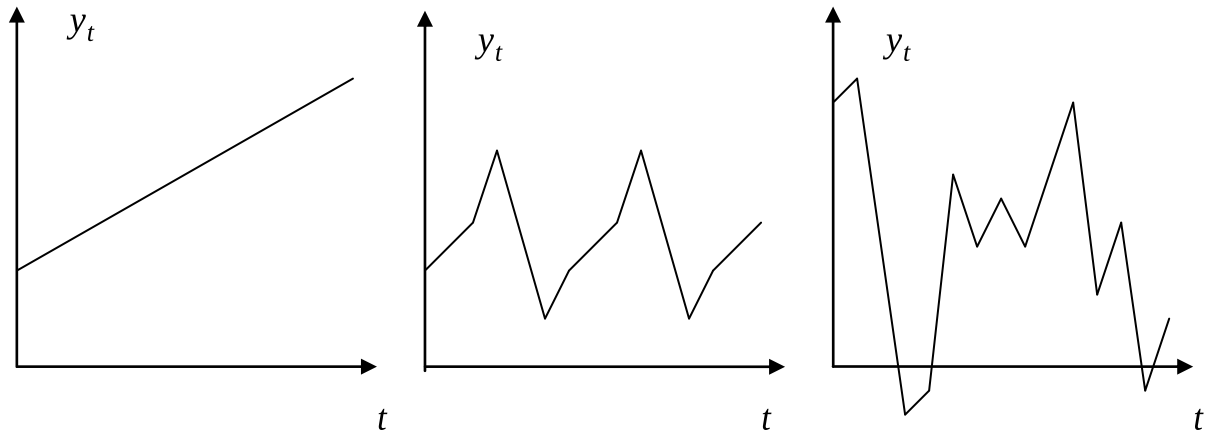


Рис. 14.1. Основные компоненты временного ряда: а) возрастающая тенденция; б) – сезонная компонента; в) – случайная компонента

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Двумя простейшими моделями, связывающими наблюдаемую величину временного ряда y_t с компонентами тренда T, цикличности C и случайности E являются модель аддитивных компонент и модель мультипликативных компонент. Модель, в которой временный ряд представлен как сумма перечисленных компонент, называется аддитивной моделью временного ряда: $Y_t = T_t + C_t + E_t$. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется мультипликативной моделью временного ряда: $Y_t = T_t \cdot C_t \cdot E_t$. Модель аддитивных компонент применима в тех случаях, когда анализируемый временной ряд имеет приблизительно одинаковые изменения на протяжении всей длительности ряда или другими словами: все значения ряда существенно убывают в пределах полосы постоянной ширины, центрированной на уровне ряда. Модель мультипликативных компонент эффективнее в тех случаях, когда изменение временной последовательности увеличивается с ростом уровня, т.е. значения ряда расходятся как имеющие тренд, а наблюдаемая последовательность значений напоминает рупор или воронку.

Основная задача эконометрического исследования отдельного временного ряда - выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

14.2. Автокорреляция уровней временного ряда и выявление его структуры

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Пример. Расчет коэффициентов автокорреляции уровней для временного ряда расходов на конечное потребление.

Пусть имеются следующие условные данные о средних расходах на конечное потребление (y_t , д.е.) за 8 лет.

Таблица 14.1

Расчет коэффициента автокорреляции первого порядка

t	y_t	y_{t-1}	$y_t - \overline{y_1}$	$y_{t-1} - \overline{y_2}$	$(y_t - \overline{y_1}) \cdot (y_{t-1} - \overline{y_2})$	$(y_t - \overline{y_1})^2$	$(y_{t-1} - \overline{y_2})^2$
1	7	-	-	-	-	-	-
2	8	7	-3,29	-3	9,87	10,8241	9
3	8	8	-3,29	-2	6,58	10,8241	4
4	10	8	-1,29	-2	2,58	1,6641	4
5	11	10	-0,29	0	0	0,0841	0
6	12	11	0,71	1	0,71	0,5041	1
7	14	12	2,71	2	5,42	7,3441	4
8	16	14	4,71	4	18,84	22,1841	16

Итого	86	70	-0,03	0	44,0	53,4287	38

$$r_1=\frac{\sum\limits_{t=2}^{n}\left(y_t-\overline{y_1}\right)\left(y_{t-1}-\overline{y_2}\right)}{\sqrt{\sum\limits_{t=2}^{n}\left(y_t-\overline{y_1}\right)^2\left(y_{t-1}-\overline{y_2}\right)^2}}=\frac{44}{\sqrt{53{,}42\cdot 38}}=0{,}976\,,$$

$$\overline{y_1}=\frac{\sum\limits_{t=2}^{n}y_t}{n-1}=\frac{8+8+10+11+12+14+16}{7}=11{,}29\,,\quad \overline{y_2}=\frac{\sum\limits_{t=2}^{n}y_{t-1}}{n-1}=\frac{7+8+8+10+11+12+14}{7}=10\,.$$

Полученное значение свидетельствует об очень тесной зависимости между расходами на конечное потребление текущего и непосредственно предшествующего годов и, следовательно, о наличии во временном ряде расходов на конечное потребление сильной линейной тенденции.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-1} (табл. 14.2).

Таблица 14.2

Расчет коэффициента автокорреляции второго порядка

t	y_t	y_{t-2}	y_t-$\overline{y_3}$	y_{t-2}-$\overline{y_4}$	$(y_t-\overline{y_3})\cdot(y_{t-2}$-$\overline{y_4})$	$(y_t-\overline{y_3})^2$	$(y_{t-2}-\overline{y_4})^2$
1	7	-	-	-	-	-	-
2	8	-	-	-	-	-	-
3	8	7	-3,83	-2,33	8,9239	14,6689	5,4289
4	10	8	-1,83	-1,33	2,4339	3,3489	1,7689
5	11	8	-0,83	-1,33	1,1039	0,6889	1,7689
6	12	10	0,17	0,67	0,1139	0,0289	0,4489
7	14	11	2,17	1,67	3,6239	4,7089	2,7889
8	16	12	4,17	2,67	11,1339	17,3889	7,1289
Итого	86	56	0,02	0,02	27,3334	40,8334	19,3334

$$r_2=\frac{\sum\limits_{t=2}^{n}\left(y_t-\overline{y_3}\right)\left(y_{t-2}-\overline{y_4}\right)}{\sqrt{\sum\limits_{t=2}^{n}\left(y_t-\overline{y_3}\right)^2\left(y_{t-2}-\overline{y_4}\right)^2}}=\frac{27{,}3334}{\sqrt{40{,}8334\cdot 19{,}3334}}=0{,}973\,,$$

$$\overline{y_3} = \frac{\sum_{t=3}^{n} y_t}{n-2} = \frac{8+10+11+12+14+16}{6} = 11{,}83, \quad \overline{y_4} = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{7+8+8+10+11+12}{6} = 9{,}33.$$

Полученные результаты подтверждают вывод о том, что ряд расходов на конечное потребление содержит линейную тенденцию.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называется лагом. Есть мнение, что максимальный лаг должен быть не больше $\frac{n}{4}$. Коэффициент автокорреляции характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию, коэффициент автокорреляции может приближаться к нулю. По знаку коэффициент автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют автокорреляционной функцией временного ряда. График зависимости ее значений от величины лага (порядка коэффициент автокорреляции) называется коррелограммой. При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

14.3. Сглаживание временных рядов с помощью скользящих средних

При анализе временного ряда существует основная задача – определения основной тенденции в развитии исследуемого явления. В некоторых случаях общая тенденция прослеживается в динамике показателя, в других ситуациях она не может просматриваться из-за существующих случайных колебаний. Например, в отдельные моменты времени сильные колебания в курсах акций могут заслонить наличие тенденции к росту или снижению этого показателя. На практике простым методом обнаружения общей тенденции является укрупнение интервалов. Например, ряд недельных данных можно преобразовать в ряд месячных данных, ряд квартальных данных – в годовые. Таким преобразованием может быть суммирование уровней исходного ряда или нахождение средних значений. Этот метод называется сглаживанием временного ряда. Суть этого метода заключается в замене фактических уровней временного ряда расчетными, которые в меньшей степени подвержены колебаниям и способствует более четкому проявлению тенденции развития.

Согласно различным подходам методы сглаживания разделяются на две группы: аналитический и алгоритмический. Аналитический подход предполагает задание общего вида функции, описывающей регулярную, неслучайную составляющую, а затем проводится статистическое оценивание неизвестных коэффициенты модели и определяются сглаженные значения уровней временного ряда путем подстановки соответствующего значения t в полученное уравнение. В алгоритмическом подходе предполагается алгоритм расчета неслучайной составляющей в любой заданный момент времени t. К алгоритмическому подходу относятся скользящие средние, которые позволяют сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии процесса.

Алгоритм сглаживания по простой скользящей средней реализуется по таким этапам. 1) Определяют длину интервала сглаживания l, включающего в себя l последовательных уровней ряда ($l < n$). Чем шире интервал сглаживания, тем в большей степени поглощаются колебания, и тенденция развития носит более плавный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания. 2) Разбивают весь период наблюдений на части и интервал сглаживания переходит по ряду с шагом равным 1. 3) Рассчитывают средние арифметические из уровней ряда, образующих каждую часть. 4) Заменяют фактические значения ряда, стоящие в центре каждой части, на соответствующие средние значения. Рекомендуют длину интервала сглаживания l выбирать нечетным числом $l = 2p + 1$, поскольку в этом случае полученные значения скользящей средней приходятся на средний член интервала. Наблюдения, которые берутся для расчета среднего значения, называются активной частью. При нечетном значении $l = 2p + 1$ все уровни активной части могут быть представлены в виде:

$$y_{t-p}, y_{t-p+1}, \ldots, y_{t-1}, y_t, y_{t+1}, \ldots, y_{t+p-1}, y_{t+p},$$

где y_t – центральный уровень активной части;

$y_{t-p}, y_{t-p+1}, \ldots, y_{t-1}$ – последовательность из p уровней активной части, предшествующей центральной;

$y_{t+1}, ..., y_{t+p-1}, y_{t+p}$ – последовательность из уровней активной части, следующей за центральной.

В этом случае скользящая средняя определяется по формуле:

$$\overline{y_t} = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + ... + y_{t+p-1}, y_{t+p}}{2p+1},$$

где y_i – фактическое значение i-го уровня;

$\overline{y_t}$ – значение скользящей средней в момент t;

$2p+1$ – длина интервала сглаживания.

При использовании простой скользящей средней выравнивание в каждой активной части проводится по прямой и аппроксимация неслучайной составляющей осуществляется с помощью линейной функции $\overline{y_t} = a_0 + a_1 t$.

Метод сглаживания приводит к устранению периодических колебаний во временном ряду, если длина интервала сглаживания равна или кратна периоду колебаний. Рекомендуется для устранения сезонных колебаний использовать скользящие средние с длиной интервала сглаживания, равной 4 или 12, однако при этом не будет выполняться условие нечетности. В этом случае скользящая средняя рассчитывается по формуле:

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + ... + y_{t-1} + y_t + y_{t+1} + ... + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} =$$

$$= \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}.$$

Поэтому для сглаживания сезонных колебаний с квартальным временным рядом или месячным временным рядом используют 4- членную и 12-членную скользящую среднюю:

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4};$$

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-6} + y_{t-51} + ... + y_t + ... + y_{t+5} + \frac{1}{2}y_{t+6}}{12}.$$

Метод простой скользящей средней рекомендуется применять, если графическое изображение временного ряда напоминает прямую. Если же данный процесс развивается нелинейно, то простая скользящая средняя приводит к существенным искажениям и в этом случае рекомендуют использовать взвешенную скользящую среднюю:

$$\overline{y_t} = \frac{\sum\limits_{i=t-p}^{t+p} y_i w_i}{\sum\limits_{i=t-p}^{t+p} w_i},$$

где w_i – весовые коэффициенты.

Весовые коэффициенты имеют свойства: 1) симметричны относительно центрального уровня; 2) сумма весов с учетом общего множителя, вынесенного за скобки, равна единице; 3) имеются как положительные, так и отрицательные веса, что позволяет сглаженной кривой сохранять различные изгибы кривой тренда.

Пример. По данным об урожайности пшеницы (y_t) за 16 лет рассчитать трехлетние и семилетние скользящие средние и графически сравнить результаты (табл. 14.3).

Таблица 14.3

Урожайность пшеницы, ц/га

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

Решение. Результаты расчетов приведены в табл. 14.4

Таблица 14.4

Расчет скользящих средних

t	y_t	$l=3$	$l=7$
1	10,3	–	–
2	14,3	10,8	–
3	7,7	12,6	–
4	15,8	12,6	13,5
5	14,4	15,6	14,9
6	16,7	15,5	15,3
7	15,3	17,4	15,3
8	20,2	17,5	15,2
9	17,1	15,0	15,5
10	7,7	13,4	16,0

11	15,3	13,1	15,8
12	16,3	17,2	15,6
13	19,9	16,9	16,1
14	14,4	17,7	–
15	18,7	17,9	–
16	20,7	–	–

При трехлетней скользящей средней:

$\bar{y}_2 = \dfrac{10{,}3+14{,}3+7{,}7}{3} = 10{,}8$; $\bar{y}_3 = \dfrac{14{,}3+7{,}7+15{,}8}{3} = 12{,}6$;

$\bar{y}_4 = \dfrac{7{,}7+15{,}8+14{,}4}{3} = 12{,}6$; $\bar{y}_5 = \dfrac{15{,}8+14{,}4+16{,}7}{3} = 15{,}6$;

$\bar{y}_6 = \dfrac{14{,}4+16{,}7+15{,}3}{3} = 15{,}5$; $\bar{y}_7 = \dfrac{16{,}7+15{,}3+20{,}2}{3} = 17{,}4$;

$\bar{y}_8 = \dfrac{15{,}3+20{,}2+17{,}1}{3} = 17{,}5$; $\bar{y}_9 = \dfrac{20{,}2+17{,}1+7{,}7}{3} = 15{,}0$;

$\bar{y}_{10} = \dfrac{17{,}1+7{,}7+15{,}3}{3} = 13{,}4$; $\bar{y}_{11} = \dfrac{7{,}7+15{,}3+16{,}3}{3} = 13{,}1$;

$\bar{y}_{12} = \dfrac{15{,}3+16{,}3+19{,}9}{3} = 17{,}2$; $\bar{y}_{13} = \dfrac{16{,}3+19{,}9+14{,}4}{3} = 16{,}9$;

$\bar{y}_{14} = \dfrac{19{,}9+14{,}4+18{,}7}{3} = 17{,}7$; $\bar{y}_{15} = \dfrac{14{,}4+18{,}7+20{,}7}{3} = 17{,}9$.

При семилетней скользящей средней:

$$\bar{y}_4 = \frac{10{,}3+14{,}3+7{,}7+15{,}8+14{,}4+16{,}7+15{,}3}{7} = 13{,}5;$$

$$\bar{y}_5 = \frac{14{,}3+7{,}7+15{,}8+14{,}4+16{,}7+15{,}3+20{,}2}{7} = 14{,}9;$$

$$\bar{y}_6 = \frac{7{,}7+15{,}8+14{,}4+16{,}7+15{,}3+20{,}2+17{,}1}{7} = 15{,}3;$$

$$\bar{y}_7 = \frac{15{,}8+14{,}4+16{,}7+15{,}3+20{,}2+17{,}1+7{,}7}{7} = 15{,}3;$$

$$\bar{y}_8 = \frac{14{,}4+16{,}7+15{,}3+20{,}2+17{,}1+7{,}7+15{,}3}{7} = 15{,}2;$$

$$\bar{y}_9 = \frac{16{,}7+15{,}3+20{,}2+17{,}1+7{,}7+15{,}3+16{,}3}{7} = 15{,}5;$$

$$\bar{y}_{10} = \frac{15,3 + 20,2 + 17,1 + 7,7 + 15,3 + 16,3 + 19,9}{7} = 16,0;$$

$$\bar{y}_{11} = \frac{20,2 + 17,1 + 7,7 + 15,3 + 16,3 + 19,9 + 14,4}{7} = 15,8;$$

$$\bar{y}_{12} = \frac{17,1 + 7,7 + 15,3 + 16,3 + 19,9 + 14,4 + 18,7}{7} = 15,6;$$

$$\bar{y}_{13} = \frac{7,7 + 15,3 + 16,3 + 19,9 + 14,4 + 18,7 + 20,7}{7} = 16,1.$$

На рис. 14.2 представлены фактический и сглаженные ряды урожайности.

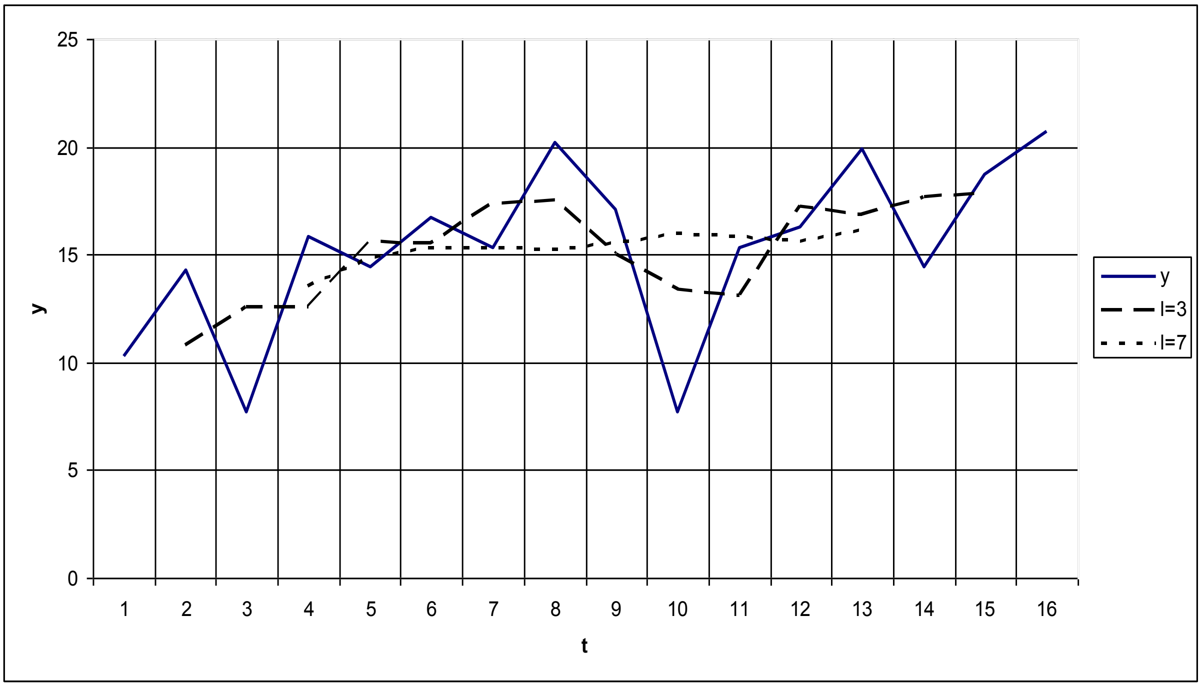


Рис. 14.2. Сглаживание ряда урожайности с помощью скользящих средних, где —— фактические уровни y_t; — — $l = 3$; ········ $l = 7$

Таким образом, ряд, сглаженный по семилетней скользящей средней ($l = 7$), имеет гладкий характер по сравнению с рядом, сглаженным по трехлетней скользящей средней ($l = 3$). Чем больше длина интервала сглаживания, тем более гладким сглаженный ряд получается.

Вопросы для самопроверки:

1) Из каких компонентов состоит временной ряд?

2) Дать определение каждой из компонент временного ряда.

3) Какая модель временного ряда называется аддитивной?

4) Какая модель временного ряда называется мультипликативной?

5) Какие существуют рекомендации о применении аддитивной и мультипликативной моделей временного ряда?
6) Что характеризует коэффициент автокорреляции?
7) Что называется автокорреляционной функцией временного ряда?
8) Что называется коррелограммой?
9) Как можно выявить структуру лага?
10) Какая основная задача анализа временного ряда?
11) Какие основные подходы в методах сглаживания?
12) Какие основные этапы алгоритма сглаживания по простой скользящей средней?

Тема 15. Моделирование тенденции временного ряда

15.1. Применение моделей кривых роста в прогнозировании основной тенденции развития

15.1. Методы выбора кривых роста и оценка адекватности и точности выбранных моделей

15.1. Применение моделей кривых роста в прогнозировании основной тенденции развития

Для моделирования тенденции развития процесса или явления в реальных экономических задачах чаще всего используют модели кривых роста. Это есть функции времени $y = f(t)$, при этом считается, что влияние других факторов несущественно или косвенного учитывается через фактор времени.

Прогнозирование на основе моделей кривых роста основывается на экстраполяции, т.е. на продление на последующие периоды тенденции, которая установлена по предыдущим периодам.

В процедуре прогнозирования на основе кривых роста выделяют такие этапы:

1) выбор одной или нескольких кривых, форма которых соответствует характеру изменения временного ряда;

2) оценка параметров выбранных кривых;

3) проверка адекватности выбранных кривых прогнозируемому процессу или явлению, оценка точности моделей и окончательный выбор кривой роста;

4) расчет точечного и интервального прогнозов.

Модели кривых роста рекомендуют разделять на три группы. К первой группе относятся функции, используемые для описания процессов с монотонным характером тенденции развития и отсутствия пределов роста. Это характерно для тенденций изменения многих экономических показателей промышленных предприятий. Ко второй группе относятся кривые, описывающий процесс, имеющий предел роста в исследуемом периоде. Такие процессы чаще всего демографические, хотя встречаются и в исследовании экономических процессов на промышленных предприятиях. Функции, относящиеся ко второму классу, называются кривыми с насыщением. Если кривые насыщения имеют точки перегиба, то они относятся к третьему классу – к S-образным кривым. По кривым третьей группы прогнозируют процессы научно-технического прогресса, нового производства продукции.

В прогнозировании экономических показателей с помощью кривых роста чаще всего применяются следующие функции:

$$y = a + bt,\ y = a + \frac{b}{t},\ y = e^{a+bt},\ y = a \cdot t^b,\ y = a + b_1 t + b_2 t^2 + ... + b_k t^k .$$

В подавляющем большинстве случаев расчет оценок параметров моделей осуществляется с помощью метода наименьших квадратов в форме регрессионных моделей, в которых в качестве зависимой переменной служат значения показателей, а фактором есть время. Для нелинейных трендовых моделей применяют процедуру линеаризации.

15.2. Методы выбора кривых роста и оценка адекватности и точности выбранных моделей

Для определения типа тенденции используют несколько способов: качественный анализ изучаемого процесса, графический метод, а также методы, использующие коэффициенты автокорреляции уровней ряда.

Наиболее простым методом выбора кривых роста считается визуальный, основный на графическом изображении временного ряда. Если на графическом изображении четко не прослеживается тенденция, то рекомендуют провести некоторые стандартные преобразования ряда (например, сглаживание) и подобрать соответствующую функцию. С помощью современных программных средств, например, статистических пакетов процедуру выбора кривой роста осуществить очень легко.

Выбор наилучшего уравнения для построения тренда можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации R^2 и выбора уравнения тренда с максимальным значением скорректированного коэффициента детерминации.

Статистическое качество построенных моделей кривых роста для прогноза проверяется по критериям проверки качества построенных регрессионных моделей: критерия Стьюдента, критерия Фишера, критерия Дарбина-Уотсона. Существование автокорреляции остатков может существенно искажать прогнозные значения.

Пример. Для проведения экономического анализа внешнеэкономической деятельности промышленного предприятия важно прогнозировать значения показателя эффективности реализации продукции на внутреннем рынке (y). В табл. 15.1 приведены квартальне значения показателя в течении 5 лет.

Таблица 15.1

Y	1,225	1,242	1,264	1,205	1,277	1,289	1,241	1,235	1,236	1,234
t	1	2	3	4	5	6	7	8	9	10
Y	1,193	1,182	1,236	1,201	1,199	1,236	1,182	1,184	1,168	1,201
t	11	12	13	14	15	16	17	18	19	20

Построить модель кривой роста показателя эффективности реализации продукции промышленного предприятия на внутреннем рынке и прогнозировать его значение на два последующих квартала.

Решение. С помощью статистического пакета Statgraphics Plus V5.1 International Professional сначала была вычислена модель кривой роста в линейной форме: $y = 1{,}2605 - 0{,}0037t$, при этом статическое качество модели было проверено с помощью критериев Стьюдента, Фишера, скорректированного коэффициента детерминации и критерия Дарбина-Уотсона: $t_a = 105{,}9$; $t_b = -3{,}74$; $F = 14{,}0$; $R^2 = 0{,}4062$; $DW = 1{,}999$. Модель относительно качественная. На последующие кварталы прогноз значения показателя эффективности реализации продукции на внутреннем рынке такой: 1,18245; 1,17873. Видим, что эффективность реализации продукции на внутреннем рынке у промышленного предприятия будет немного снижаться в последующих двух кварталах.

В табл. 15.2 представлены рассчитанные альтернативные модели по основным кривым роста по рейтингу значений коэффициентов детерминации.

Таблица 15.2

Альтернативные модели	Скорректированный коэффициент детерминации
$y = \dfrac{1}{0,7931 + 0,0025t}$	0,4124
$y = e^{0,2317 - 0,003t}$	0,4094
$y = (,1228 - 0,0017t)^2$	0,4078
$y = 1,2605 - 0,0037t$	0,4062
$y = 1,248 - 0,00029t - 0,000163t^2$	0,3961
$y = 1,2829 - 0,0199\sqrt{t}$	0,3407
$y = 1,2672t^{-0,0175}$	0,2348
$y = 1,2667 - 0,0213 ln t$	0,2320
$y = \dfrac{1}{0,8242 - \dfrac{0,0275}{t}}$	0,0256
$y = e^{0,1937 + \frac{0,033}{7}}$	0,0238
$y = 1,2142 + \dfrac{0,0403}{t}$	0,0219

С табл.2 видим, что исходя из значения скорректированного коэффициента детерминации наилучшей моделью есть $y = \dfrac{1}{0,7931 + 0,0025t}$. Прогноз по этой модели эффективности реализации продукции на внутреннем рынке такой: 1,18282; 1, 17935. Расхождение прогнозных значений, рассчитанных по двум моделям различаются в тысячных единицах.

Специалисты по статистическим методам прогнозирования считают, что не может быть чисто формальных подходов к выбору методов и моделей прогнозирования. Успешное применение статистических методов прогнозирования на практике возможно лишь при сочетании знаний в области самих методов с глубоким знанием объекта исследования, с содержательным анализом изучаемого явления.

Вопросы для самопроверки:

1) Из каких этапов состоит процедура прогнозирования на основе кривых роста?
2) Какие типы модели кривых роста различаю?
3) Какие функции чаще всего используются для прогнозирования экономических показателей?
4) Перечислить методы выбора кривых роста.
5) Как осуществить выбор наилучшего уравнения для построения тренда?

Тема 16. Моделирование сезонных и циклических колебаний

16.1. Общая характеристика методов моделирования сезонных и циклических колебаний

16.2. Статистические методы оценки уровня сезонности

16.1. Общая характеристика методов моделирования сезонных и циклических колебаний

Проблемами периодических колебаний в экономике занимаются давно. Известные ученые К. Жюгляр, С. Китчин, С. Кузнец, Н. Кондратьев, позднее У. Персонс, У. Митчелл, Е.Е. Слуцкий, современники – Четвериков, Ферстер, Шискин, Эйзенпресс внесли существенный вклад в решение этих проблем .

Как известно, моделирование циклических колебаний производится почти по той же схеме, что и моделирование сезонных колебаний, хотя моделирование циклов с длинными периодами имеет свои особенности [с. 161 – 167].

При моделировании сезонных и циклических колебаний используют несколько подходов:

1) на основе метода скользящих средних;

2) с применением *dummy*-переменных;

3) гармонический анализ;

4) на основе индексов сезонности;

5) сезонные разностные операторы;

6) адаптивные модели.

Модель сезонных колебаний с применением *dummy*-переменных рассмотрена в теме 10 «Обобщенные схемы регрессионного анализа». Следует отметить, что число *dummy*-переменных должно быть на единицу меньше числа различных периодов времени внутри одного сезонного цикла (при моделировании поквартальных колебаний таких переменных три, помесячных - одиннадцать). Каждая *dummy*-переменная равна единице для определенного периода и нулю для остальных.

Моделирование сезонных колебаний методами гармонического анализа основывается на использовании гармоник ряда Фурье:

$$y_t = a_0 + \sum_{k=1}^{m} (a_k \cos kt + b_k \sin kt),$$

где k - номер гармоники (обычно целое число от 1 до 4).

Параметры данного уравнения определяются при помощи МНК:

$$a_0 = \frac{\sum_t y_t}{n}; \quad a_t = \frac{2}{n}\sum_t y_t \cos kt; \quad b_t = \frac{2}{n}\sum_t y_t \sin kt \ .$$

Значения t изменяются от 0 с приростом $\frac{2\pi}{n}$, $n -$ число уравнений временного ряда внутри сезонного цикла.

Расчет индексов сезонности основывается на нескольких подходах, исходя из динамики исследуемых процессов.

Если среднегодовые значения уровня ряда на протяжении исследуемого периода остается сравнительно неизменным, то индексы сезонности вычисляют по формулам (способ постоянной средней):

$$i_t = \frac{\overline{y_k}}{\overline{\overline{y}}} \cdot 100,$$

где $\overline{y_k}$ - средняя арифметическая фактических уровней одновременных месяцев (кварталов);

$k = \overline{1,12}$ - для месяцев, $k = \overline{1,4}$ - для кварталов;

$\overline{y}$ - средняя арифметическая уровней за исследуемый период.

Если в течение каждого года исследуемого периода повышающийся или снижающийся тренд отсутствует или не значителен, то индексы сезонности могут быть вычислены по формуле:

$$i_t = \frac{\sum_{l=1}^{m} \frac{y_{lk}}{\overline{y_l}}}{m},$$

где y_{lk} - уровень временного ряда для k-го месяца (квартала) l -го года;

$\overline{y_l}$ - средняя арифметическая фактических уровней ряда l -го года;

m - число лет в исследуемом периоде.

Применение сезонных разносных операторов предназначено для исключения сезонных компонент из временного ряда и предполагает переход от исходного временного ряда к ряду $y_t, i=\overline{1,n}$ к ряду $\Delta y_t = y_t - y_{t-\tau}, i=\tau+1,...,n$, где τ - период сезонности. Иногда используют сезонные разностные операторы более высоких порядков, например, второго порядка $\Delta^2 y_t = y_t - y_{t-\tau}$.

Адаптивные методы и модели учитывают информационную неравноценность исходных данных, например, при прогнозировании ретроспективная информация не равнозначна и следует больше учитывать на данные, расположенные ближе к настоящему моменту времени.

У истоков адаптивных методов лежит модель экспоненциального сглаживания.

Адаптивные модели приспосабливаются к новой информации, отражающей изменения условий. К адаптивным моделям относятся модели Брауна, Хольта, Хольта-Уинтерса, модели авторегрессии и другие.

Рекомендуется следующая схема построения адаптивной модели :

1) по нескольким первым уравнениям временного ряда находят оценки параметров модели;

2) по построенной модели находят прогноз на один шаг вперед;

3) вычисляют ошибку прогнозирования;

4) корректируют модели с учетом найденной ошибки;

5) по скорректированной модели рассчитывается прогноз еще на один шаг и так далее.

Отмечается важнейшее достоинство адаптивных методов – построение самокорректирующихся моделей, способных учитывать результат прогноза, сделанного на предыдущем шаге. Поэтому адаптивные методы особенно удачно используются при оперативном, краткосрочном прогнозировании.

16.2. Статистические методы оценки уровня сезонности

В настоящее время при описании и прогнозировании тренд-сезонных процессов используются комбинированные подходы, связанные с применением индексов сезонности вместе кривыми роста, процедуры, опирающиеся на адаптивные модели, сезонный вариант модели ARIMA, специализированные подходы, учитывающие особенности конкретных временных рядов.

Пусть динамика ряда характеризуется неизменными во времени сезонными эффектами. Процедура расчета сезонной составляющей зависит от принятой модели временного ряда – аддитивной или мультипликативной. Выбор формы модели зависит от структуры сезонных колебаний. Если амплитуда колебаний почти постоянна, то разрабатывают аддитивную модель временного ряда, в которой значения сезонной компоненты считаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, то разрабатывают мультипликативную модель временного ряда. В аддитивной модели характеристики сезонности измеряются в абсолютных величинах, в мультипликативной – в относительных.

Алгоритм расчета для аддитивной сезонности состоит из четырех этапов [5, с. 84 – 87]. На первом этапе предполагается описание тенденции с помощью скользящей средней при четной длине интервалов сглаживания $l = 2p$. Скользящая средняя для временных рядов месячной динамики определяется по формуле:

$$y_t^{'} = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \ldots + y_t + \ldots + y_{t+5} + \frac{1}{2}y_{t+6}}{12};$$

для рядов квартальной динамики:

$$y_t^{'} = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

На втором этапе рассчитывается отклонение фактических значений от уровней сглаженного ряда: $x_t = y_t - y_t^{'}$. Уровни вновь полученного ряда отражают эффект сезонности и случайности.

На третьем этапе для элиминирования влияния случайных факторов определяется предварительные значения сезонной составляющей как средние значения из уровней x_t для одноименных месяцев (кварталов).

На четвертом этапе проводится корректировка первоначальных значений сезонной составляющей. Для аддитивного случая сумма значений сезонной составляющей для полного сезонного цикла равняется нулю. Поэтому скорректированные оценки сезонной компоненты определяются так: $S_i = \overline{x_i} - \bar{x}$, где $\bar{x} = \frac{1}{m}\sum_{i=1}^{m}\overline{x_i}$, m- число фаз в сезонном цикле (для месячной динамики $m = 12$, для квартальных данных $m = 4$).

В алгоритм расчета для мультипликативной сезонности меняется второй и четвертой этапы, а первый и третий остаются без изменения. На втором этапе рассчитывается отклонение фактических значений от уровней сглаженного ряда как $x_t = \frac{y_t}{y_t'}$. Взаимопогашаемость сезонных колебаний в мультипликативной форме выражается в том, что средняя арифметическая значений коэффициентов сезонности для сезонного цикла равняется единице. Окончательные оценки коэффициентов сезонности будут: $S_i = \overline{x_i}\,\bar{x}, \quad i = \overline{1,m}$, где $\bar{x} = \frac{m}{\sum_{i=1}^{m}\overline{x_i}}$, m- число фаз в сезонном цикле.

Процедуру построения тренд-сезонных моделей рекомендуют проводить в следующей последовательности.

1) Оценивание сезонной составляющей проводить по рассмотренным алгоритмам с учетом характера сезонности: аддитивной или мультипликативной.

2) Сезонная корректировка исходных данных.

3) Расчет параметров тренда на основе временного ряда, полученного на предыдущем шаге.

4) Моделирование динамики исходного ряда с учетом трендовой и сезонной составляющих.

5) Оценка точности и адекватности полученной модели.

6) Применение вычисленной модели для прогнозирования.

Вопросы для самопроверки:

1) Перечислить подходы в моделировании сезонных и циклических колебаний.

2) В чем состоит отличие каждого из подходов в моделировании сезонных и циклических колебаний.
3) От чего зависит процедура расчета сезонной составляющей?
4) Привести этапы алгоритма расчета для аддитивной сезонности.
5) В чем состоит отличие алгоритма расчета для мультипликативной сезонности?
6) Раскрыть содержание процедуры построения тренд-сезонных моделей.

Использованная литература

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. – М.: ЮНИТИ, 1998с. – 1022 с.
2. Бородич С. А. Эконометрика : учеб. пособие / С. А. Бородич. – Мн. : Новое знание, 2001. – 408 с.
3. Вильямс Дж. Совершенный стратег или Букварь по теории стратегических игр. Пер. с англ. –М., «Советское радио», 1960. –270 с.
4. Доугерти К. Введение в эконометрику / К. Доугерти; пер. с англ. – М. : ИНФРА-М, 1999. – 402 с.
5. Дуброва Т.А. Статистические методы прогнозирования: Учеб. пособие для вузов. – М. : ЮНИТИ-ДАНА, 2003. – 206 с.
6. Єгоршин О. О. Тексти лекцій з курсу “Математичне програмування” курсу “Математика для економістів.” Ч. ІІІ./ О. О. Єгоршин, Л. М. Малярець. – Харків : Вид. ХДЕУ, 2001 – 88 с.
7. Єгоршин О. О. Математичне програмування : [підруч. для студ. вищ. навч. закл.] / О. О. Єгоршин, Л. М. Малярець. – Харків : ВД «ІНЖЕК», 2006. – 384 с.
8. Егоршин А. А. Корреляционно-регрессионный аналіз : [учеб. пособ. для студ. экон. спец. вузов] / А. А. Егоршин, Л. М. Малярец. – Харків, : Основа, 1998. – 208 с.
9. Егоршин А. А. Математическое программирование : учебное пособие / А. А. Егоршин, Л. М. Малярец. – Харьков : ИД «ИНЖЭК», 2003. – 240 с.
10. Єгоршин О.О. Лабораторний практикум з навчальної дисципліни «Економіко-математичні методи та моделі: економетрика»: навчально-практичний посібник / О. О. Єгоршин, Л. М. Малярець. – Х.: Вид. ХНЕУ, 2011. – 148 с.
11. Елисеева И. И. Практикум по эконометрике : учебн. пособ. / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко ; [под ред.

И. И. Елисеевой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2006. – 344 с.

12. Кузнецов Ю. Н. Математическое программирование : учеб. пособие / Ю. Н. Кузнецов, В. И. Кузубов, А. Б. Волощенко. – 2-е изд., перераб. и доп.– М. : Высшая школа, 1980 – 300 с.
13. Лабораторний практикум з навчальної дисципліни «Економіко-математичне моделювання» : навчально-практичний посібник/ Л. М. Малярець, П. М. Куликов, І. Л. Лебедєва, Л. О. Норик, О. Г. Тижненко. – Харків : Вид. ХНЕУ, 2009. – 136 с.
14. Лугін О. Є. Економетрія : навчальний посібник / О. Є. Лугін, С. В. Білоусова, О. М. Білоусов. – К. : Центр навчальної літератури, 2005. – 252 с.
15. Лук'яненко І. Г. Економетрика : підручник. / І. Г. Лук'яненко, Л. І. Краснікова. – К. : Тов. «Знання», КОО, 1998. – 494 с.
16. Магнус Я. Р. Эконометрика : начальный курс. / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 1997. – 248 с.
17. Маленво Э. Статистические методы эконометрии. Вып. 1 / Э. Маленво. – М. : Статистика, 1975. – 424 с.
18. Маленво Э. Статистические методы эконометрии. Вып. 2 / Э. Маленво. – М. : Статистика, 1975. – 325 с.
19. Малярець Л.М. Економіко-математичне моделювання: навчальний посібник. Харків: Вид. ХНЕУ, 2010. – 312 с.
20. Математичні методи в сучасних економічних дослідженнях: монографія. За заг. ред.. д.е.н., проф.. Малярець Л.М. Харків: Вид. ХНЕУ, 2011. – 272 с.
21. Назаренко О. М. Основи економетрики : підручник. / О. М. Назаренко. – К.: Центр навчальної літератури, 2004. – 392 с.
22. Наконечний С. І. Економетрія : підручник / Наконечний С. І., Терещенко Т. О., Романюк Т. П. – 3-те вид., доп. та перероб. – К. : КНЕУ, 2005. – 520 с.
23. Наконечний С. І. Математичне програмування : навч. посібник / Наконечний С. І., Савіна С. С. – К. : КНЕУ, 2005. – 452 с.
24. Пономаренко В. С. Багатовимірний аналіз соціально-економічних систем : навчальний посібник / Пономаренко В. С., Малярець Л. М. – Харків : Вид. ХНЕУ, 2009. – 384 с.
25. Практикум по эконометрике : учеб. пособие / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко и др. ; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2007. – 344 с.

26. Просветов Г.И. Эконометрика. Задачи и решения: Учебно-методическое пособие. – М.: Издательство РДЛ, 2004. – 104 с.
27. Таха Хемди А. Введение в исследование операций / Таха Хемди А.; пер. с англ. 7-е изд. – М. : Издательский дом «Вильямс», 2005. – 912 с.
28. Тихомиров Н. П. Эконометрика : учебник / Тихомиров Н. П., Дорохина Е. Ю. – М. : Изд. «Экзамен», 2003. – 512 с.
29. Христиановский В. В. Прикладная эконометрия : [учебн. для экон. вузов] / В. В. Христиановский, Н. Г. Гузь, О. Г. Кривенчуг. – Донецк : ДонГУ, 1998. – 172 с.
30. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа: Руководство для экономистов/ Пер. с нем. и предисл. В.М. Ивановой. – М. Финансы и статистика, 1983. – 302 с.
31. Фролькис В. А. Введение в теорию и методы оптимизации для экономистов / Фролькис В. А. – 2-е изд. – СПб. : Питер, 2002. – 320 с.
32. Эконометрика : учебник / под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2004. – 344 с.
33. Экономико-математические методы и модели : [учеб. пособ.] / под общ. ред. А. В. Кузнецова. – Мн. : БГЭУ, 1999. – 416 с.
34. Экономико-математические методы и прикладные модели : учеб. пособие для вузов / под ред. В. В. Федосеева. – М. : ЮНИТИ, 1999. – 392 с.
35. Экономико-математический энциклопедический словарь / [гл. ред. В. И. Данилов-Данильян]. – М. : Большая Российская энциклопедия: Изд. дом «ИНФРА-М», 2003. – 688 с.
36. Эзекиэл М, К.А. Фокс. Методы корреляции и регрессии. – М : Статистика, 1966. – 324 с.
37. **Экономико-математический словарь : словарь современной экономической науки / [авт.-состав. Лопатников Л. И.] – М. : Изд. «ABF», 1996. – 704 с.**

Содержание