ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ЭКОНОМИКИ И УПРАВЛЕНИЯ

А.Т. Семенов,
Н.В. Воронович

ЭКОНОМЕТРИКА

Учебно-методический комплекс
для заочной и дистанционной форм обучения

Новосибирск
2006

ББК

С

Издается в соответствии с планом учебно-методической работы НГУЭУ

С 00 **Семенов А.Т., Воронович Н.В.**

Эконометрика: Учебно-методический комплекс. – Новосибирск: НГУЭУ, 2006. – 108 с.

Учебно-методический комплекс по эконометрике предназначен для студентов НГУЭУ заочной и дистанционной форм обучения. В него входят предусмотренные стандартом компоненты: рабочая программа, методические рекомендации, курс лекций и другие материалы. Основу комплекса составляет курс лекций, знакомящий студентов с ключевыми разделами эконометрики. Комплекс также включает необходимые при изучении и подготовке к экзамену словарь терминов, список основной и дополнительной литературы, контрольные вопросы, задания для самостоятельной работы по ключевым разделам дисциплины.

Представленные в комплексе материалы могут быть использованы также студентами очной формы обучения.

ББК

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ «ЭКОНОМЕТРИКА»

Раздел 1. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЙ

1.1. ВЫПИСКА ИЗ ГОСУДАРСТВЕННЫХ ОБРАЗОВАТЕЛЬНЫХ СТАНДАРТОВ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

Линейная модель множественной регрессии.

Метод наименьших квадратов (МНК); свойства оценок МНК.

Показатели качества регрессии.

Линейные регрессионные модели с гетероскедастичными и автокоррелированными остатками.

Обобщенный метод наименьших квадратов (ОМНК).

Регрессионные модели с переменной структурой (фиктивные переменные).

Нелинейные модели регрессии и их линеаризация.

Характеристики временных рядов.

Модели стационарных и нестационарных временных рядов, их идентификация.

Система линейных одновременных уравнений.

Косвенный, двухшаговый и трехшаговый метод наименьших квадратов.

1.2. ЦЕЛИ И ЗАДАЧИ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью дисциплины «Эконометрика» является освоение студентами методов исследования взаимосвязей экономических переменных на основе построения и анализа эконометрических моделей; овладение навыками решения конкретных задач по выявлению, оценке и анализу количественных зависимостей между различными показателями экономических объектов и процессов; формирование умения вырабатывать практические рекомендации на основе результатов эконометрического исследования.

В соответствии с целью дисциплины ее основными задачами являются:

• освоение методов выявления, оценки и анализа сложных взаимосвязей между экономическими показателями;

• формирование навыков построения эконометрических моделей, овладение методами нахождения оценок неизвестных параметров моделей;

• овладение методами обработки и подготовки исходной статистической информации для проведения расчетов по эконометрическим моделям;

• приобретение навыков использования компьютерных технологий при исследовании экономических объектов и процессов с помощью эконометрических моделей;

• формирование навыков разработки прогнозов для исследуемых экономических показателей, выработки практических рекомендаций на основе результатов, полученных при расчетах по эконометрическим моделям.

1.3. ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ ДИСЦИПЛИНЫ

По окончании изучения дисциплины «Эконометрика» слушатель должен:

• ***иметь представление*** о методах эконометрического моделирования, основных эконометрических моделях, методах оценивания неизвестных параметров моделей, проверки надежности найденных оценок и модели в целом; об уравнениях парной и множественной регрессии, системах одновременных уравнений, методах прогнозирования; о проблемах, возникающих в эконометрическом моделировании, и способах их решения, о возможностях применения компьютерных технологий для выявления, изучения и анализа сложных взаимосвязей между экономическими переменными;

• ***знать*** основные понятия дисциплины: взаимосвязи между экономическими показателями, эконометрическая модель, этапы эконометрического моделирования, оценки неизвестных параметров, уравнение регрессии, коэффициенты парной, множественной, частной и ранговой корреляции, надежность коэффициентов уравнения регрессии, значимость коэффициентов корреляции, надежность (адекватность) модели, доверительный интервал, точечный и интервальный прогноз, мультиколлинеарность, автокорреляция, гетероскедастичность, временной ряд, трендовая модель; основные эконометрические модели: линейная с двумя переменными, общая линейная модель, нелинейные модели, трендовые модели; методы нахождения оценок неизвестных параметров моделей: метод наименьших квадратов, косвенный метод наименьших квадратов, двухшаговый метод наименьших квадратов; критерии проверки надежности коэффициентов уравнений регрессии, значимости коэффициентов корреляции, адекватности модели; методы и способы проведения расчетов на персональном компьютере по эконометрическим моделям;

• ***уметь*** выявлять наличие или отсутствие взаимосвязей между экономическими переменными, выдвигать гипотезы о виде статистической зависимости между исследуемыми показателями, формулировать математические модели исследуемой взаимосвязи; применять метод наименьших квадратов для получения оценок неизвестных параметров моделей, проверять надежность полученных оценок, содержательно интерпретировать значения коэффициентов уравнения регрессии; вычислять коэффициенты парной, множественной и частной корреляции, коэффициент детерминации, проверять их значимость, содержательно интерпретировать их численные значения; находить оценки неизвестных параметров нелинейных моделей, выбирать из нескольких нелинейных моделей наиболее адекватную; находить точечные и интервальные прогнозы исследуемой зависимой переменной; выявлять наличие или отсутствие мультиколлинеарности независимых переменных, автокорреляции остатков, гетероскедастичности случайных возмущений, устранять или смягчать проблемы мультиколлинеарности, автокорреляции и гетероскедастичности; идентифицировать системы одновременных уравнений, применять косвенный и двухшаговый метод наименьших квадратов для оценки неизвестных параметров систем одновременных уравнений; исследовать временные ряды, строить трендовые модели, оценивать неизвестные параметры трендовых моделей, строить точечные и интервальные прогнозы временных показателей; проводить разнообразные расчеты на персональном компьютере по эконометрическим моделям.

1.4. ФОРМЫ КОНТРОЛЯ

Итоговый контроль. Для контроля усвоения дисциплины «Эконометрика» учебными планами предусмотрен *экзамен* или *зачет*. Экзаменационная оценка является итоговой по дисциплине и проставляется в приложении к диплому (выписке из зачетной книжки).

Текущий контроль. Студенты заочной и дистанционной форм обучения выполняют одну контрольную работу по данной дисциплине. Результат выполнения контрольной работы является основанием для оценки текущего контроля. Студенты, не выполнившие контрольную работу или не получившие по ней «зачет», не допускаются к экзамену, как не выполнившие график учебного процесса по данной дисциплине.

Раздел 2. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2.1. ТЕМАТИЧЕСКИЙ ПЛАН УЧЕБНОЙ ДИСЦИПЛИНЫ (РАСПРЕДЕЛЕНИЕ ЧАСОВ)

Наименование тем	Количество часов		
	Лекции	Самостоятельная работа	Всего часов по теме
Тема 1. Введение в эконометрику	0,5	1,5	2
Тема 2. Линейная эконометрическая модель с двумя переменными	3	24	27
Тема 3. Общая линейная эконометрическая модель	2,5	24,5	27
Тема 4. Нелинейные эконометрические модели	1	15	16
Тема 5. Проверка выполнения основных предположений регрессионного анализа	2	14	16
Тема 6. Системы одновременных уравнений	1	19	20
Тема 7. Анализ временных рядов	2	26	28
Итого по дисциплине	12	124	136

2.2. СОДЕРЖАНИЕ ОТДЕЛЬНЫХ ТЕМ

Тема 1. Введение в эконометрику

Эконометрика как наука и ее связь с другими дисциплинами. Понятие эконометрического моделирования и его основные этапы.

Тема 2. Линейная эконометрическая модель с двумя переменными

Взаимосвязи между экономическими переменными. Модель парной линейной регрессии; корреляционное поле (диаграмма рассеяния); оценка неизвестных параметров модели по методу наименьших квадратов (МНК). Предпосылки парного линейного регрессионного анализа. Свойства МНК-оценок. Интервальное оценивание параметров модели, функции регрессии и индивидуальных значений результирующей переменной. Проверка качества парной регрессионной модели: оценка значимости параметров модели и уравнения регрессии. Коэффициент детерминации и его смысл. Прогнозирование на основе уравнения регрессии. Коэффициент парной корреляции: определение, свойства, содержательный смысл; проверка значимости.

Тема 3. Общая линейная эконометрическая модель

Линейная модель множественной регрессии и оценка ее параметров методом наименьших квадратов. Предпосылки множественного линейного регрессионного анализа. Свойства МНК-оценок. Интервальное оценивание параметров модели, функции регрессии и индивидуальных значений результирующей переменной. Анализ качества модели множественной линейной регрессионной модели: оценка значимости параметров модели и уравнения регрессии. Коэффициент детерминации и его смысл. Прогнозирование на основе множественной регрессионной модели. Частная корреляция.

Тема 4. Нелинейные эконометрические модели

Нелинейные модели регрессии и их линеаризация. Применение МНК для оценки параметров нелинейных моделей. Производственная функция Кобба-Дугласа. Методы определения наиболее адекватной нелинейной модели: коэффициент детерминации и средняя ошибка аппроксимации.

Тема 5. Проверка выполнения основных предположений регрессионного анализа

Гетероскедастичность: суть, последствия, методы обнаружения и устранения гетероскедастичности.

Автокорреляция остатков: суть и причины, последствия автокорреляции, методы ее обнаружения и устранения.

Мультиколлинеарность: суть и последствия мультиколлинеарности, методы определения и устранения.

Тема 6. Системы одновременных уравнений

Системы эконометрических уравнений. Виды систем эконометрических уравнений: системы независимых, рекурсивных и взаимозависимых уравнений. Понятия эндогенных, экзогенных и предопределенных переменных в системах эконометрических уравнений. Структурная и приведенная форма модели. Модель Кейнса формирования доходов и модель «спроса-предложения».

Идентифицируемость уравнений. Условие идентифицируемости уравнений, необходимые и достаточные условия идентификации.

Идентифицируемые уравнения. Косвенный метод наименьших квадратов.

Сверхидентифицируемые уравнения. Двухшаговый метод наименьших квадратов.

Тема 7. Анализ временных рядов

Задачи анализа временных рядов: определение временных рядов и особенности их анализа, компоненты временного ряда. Сглаживание (выравнивание) временных рядов: аналитические и алгоритмические методы. Трендовые модели временных рядов: метод наименьших квадратов для оценки параметров линейной модели, методы оценивания неизвестных параметров нелинейных трендов. Методы учета сезонности при построении и исследовании трендов. Прогнозирование на основе трендовых моделей временных рядов.

Раздел 3. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

3.1. СПИСОК ВОПРОСОВ ДЛЯ ПОДГОТОВКИ К ЭКЗАМЕНУ

1. Взаимосвязи между экономическими переменными, понятие «эконометрика», основные задачи эконометрики, основные этапы эконометрического моделирования.
2. Линейная модель с двумя переменными: постановка задачи, формулировка модели, необходимость включения в модель случайной составляющей, предположения относительно случайной составляющей модели.
3. Линейная модель с двумя переменными: оценка неизвестных параметров методом наименьших квадратов, система нормальных уравнений, уравнение регрессии, содержательная интерпретация коэффициентов уравнения, график функции регрессии.
4. Свойства оценок параметров линейной модели с двумя переменными: несмещенность, состоятельность и эффективность. Остаточная дисперсия уравнения парной регрессии и стандартные ошибки коэффициентов регрессии.
5. Интервальное оценивание коэффициентов регрессии, среднего и индивидуального значений зависимой переменной парной регрессии: постановка задачи и формулы доверительных интервалов для теоретических коэффициентов регрессии, среднего и индивидуального значений зависимой переменной.
6. Проверка качества подбора уравнения парной регрессии (адекватности модели): значимость коэффициентов регрессии и уравнения регрессии; коэффициент детерминации и его смысл; критерий Фишера.
7. Прогнозирование на основе регрессионной модели: прогноз среднего и индивидуального значений зависимой переменной.

8. Коэффициент парной корреляции: определение, свойства, содержательный смысл, проверка значимости. Связь с коэффициентами регрессии и детерминации.

9. Общая линейная регрессионная модель: постановка задачи, формулировка модели, предположения, оценка неизвестных параметров методом наименьших квадратов, система нормальных уравнений, уравнение множественной регрессии, его содержательный смысл.

10. Свойства оценок общей линейной модели: несмещенность, состоятельность и эффективность. Остаточная дисперсия уравнения множественной регрессии и стандартные ошибки коэффициентов регрессии.

11. Интервальное оценивание коэффициентов регрессии, среднего и индивидуального значений зависимой переменной множественной регрессии: постановка задачи и формулы доверительных интервалов для теоретических коэффициентов регрессии, среднего и индивидуального значений зависимой переменной.

12. Проверка качества подбора уравнения множественной регрессии (адекватности модели): значимость коэффициентов регрессии и уравнения регрессии; коэффициент детерминации и его смысл; критерий Фишера.

13. Прогнозирование на основе множественной регрессионной модели: прогноз среднего и индивидуального значений зависимой переменной.

14. Коэффициенты частной корреляции. Связь коэффициентов парной и частной корреляции.

15. Нелинейные эконометрические модели: постановка задачи, формулировка модели. Основные типы нелинейных моделей: полиномиальная, степенная, показательная, гиперболическая. Методы приведения нелинейных моделей к линейным.

16. Нелинейные эконометрические модели: оценка неизвестных параметров методом наименьших квадратов, методы выбора наиболее адекватной модели.

17. Производственная функция Кобба-Дугласа: свойства функции, примеры.

18. Гетероскедастичность: суть и последствия гетероскедастичности, методы обнаружения и устранения.

19. Автокорреляция остатков: суть и последствия автокорреляции, методы обнаружения и устранения.

20. Мультиколлинеарность: суть и последствия мультиколлинеарности, методы определения и устранения мультиколлинеарности.

21. Системы одновременных уравнений. Виды систем эконометрических уравнений: независимые, рекурсивные, взаимозависимые. Структурная и приведенная форма модели. Модель Кейнса формирования доходов и модель «спроса-предложения».

22. Идентифицируемость уравнений. Необходимые и достаточные условия идентифицируемости.

23. Идентифицируемые уравнения. Косвенный метод наименьших квадратов.

24. Сверхидентифицируемые уравнения. Двухшаговый метод наименьших квадратов.

25. Определение и компоненты временных рядов; задачи анализа временных рядов.

26. Сглаживание временных рядов: аналитические и алгоритмические методы.

27. Трендовые модели временных рядов: метод наименьших квадратов для оценки параметров линейной модели, прогнозирование на основе линейных трендов.

28. Нелинейные трендовые модели: метод наименьших квадратов для оценки неизвестных параметров нелинейных трендов, выбор наиболее адекватной модели, прогнозирование.

29. Методы учета сезонных колебаний при построении, анализе и применении трендовых моделей.

3.2. СПИСОК ЛИТЕРАТУРЫ

Основная:

1. *Воронович Н.В.*, *Русин Г.Л.* Эконометрика: Методические указания по выполнению контрольных работ. – Новосибирск, НГУЭУ, 2005.
2. *Доугерти К.* Введение в эконометрику. – М.: ИНФРА-М, 1999.
3. *Практикум* по эконометрике / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001.
4. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002.

Дополнительная:

1. *Айвазян С.А.*, *Мхитарян В.С.* Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.
2. *Афанасьев В.Н.*, *Юзбашев Н.Н.* Анализ временных рядов и прогнозирование. – М.: Финансы и статистика, 2001.
3. *Бородич С.А.* Эконометрика. – Минск: Новое Знание, 2001.
4. *Джонстон Дж.* Эконометрические методы. – М.: Статистика, 1980.
5. *Кремер Н.Ш.*, *Путко Б.А.* Эконометрика. – М.: ЮНИТИ-ДАНА, 2002.
6. *Магнус Я.Р.*, *Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс. – М.: Дело, 2000.
7. *Семенов А.Т.* Таблицы вероятностных распределений и квантилей: Учебное пособие. – Новосибирск, НГАЭиУ, 1998.
8. *Семенов А.Т.* Теория вероятностей и математическая статистика: Учебно-методический комплекс. – 2-е изд., перераб. и доп. – Новосибирск: НГУЭУ, 2004.

ТЕКСТЫ ЛЕКЦИЙ
УЧЕБНОЙ ДИСЦИПЛИНЫ
«ЭКОНОМЕТРИКА»

ПРЕДИСЛОВИЕ

Дисциплина «Эконометрика», наряду с микро- и макроэкономикой, входит в число базовых дисциплин в системе высшего экономического образования. Это и неудивительно, поскольку современные экономические теории и исследования требуют достаточно хорошего владения математическим аппаратом для изучения взаимосвязей экономических явлений и процессов.

При изложении учебного материала предполагается, что студенты, изучающие эконометрику, уже прослушали базовые курсы по высшей математике, теории вероятностей и математической статистике в объеме экономического вуза.

В *теме* 1 изложены основные аспекты эконометрического моделирования, его предпосылки, виды моделей, этапы проведения и возникающие при этом проблемы.

В *темах* 2, 3 рассмотрены классические линейные регрессионные модели: в *теме* 2 – парные регрессионные модели, на примере которых удается наиболее доступно и наглядно продемонстрировать основные понятия регрессионного и корреляционного анализа; в *теме* 3 – обобщение регрессии на случай нескольких объясняющих переменных. На примере парной линейной регрессии представлен фундаментальный метод оценки параметров уравнений регрессии – метод наименьших квадратов.

Четвертая тема посвящена нелинейным регрессионным моделям, часто используемым для описания взаимосвязей экономических показателей. Приводятся примеры, методы линеаризации и оценивания моделей.

В *теме* 5 рассмотрен ряд проблем, связанных с нарушением основных предпосылок регрессионного анализа – гетероскедастичностью и автокоррелированностью остатков, их обнаружением и устранением, а также с мультиколлинеарностью между объясняющими переменными в модели множественной линейной регрессии.

В *шестой теме* анализируются эконометрические модели, представленные системой одновременных уравнений. Выясняются причины невозможности использования стандартных методов, характерных для индивидуальных уравнений, и рассматриваются методы нахождения оценок для систем одновременных уравнений.

В *теме* 7 даны общие понятия, связанные с временными рядами и использованием моделей на основе временных рядов для прогнозирования.

Изложение материала для большей наглядности сопровождается иллюстрирующими его примерами и задачами с решениями. Каждая тема завершается контрольными вопросами для самопроверки и заданиями для самостоятельного решения. В конце комплекса в Приложении приведены таблицы, необходимые для практических расчетов по излагаемой в пособии методике.

Параграфы в темах, формулы, примеры и замечания имеют двойную нумерацию: первая цифра означает номер темы, вторая – номер соответствующего объекта в теме.

Для удобства читателей основные понятия, формулировки результатов и утверждений выделены *курсивом*. Окончания примеров и замечаний отмечены в тексте соответственно значками ■ и ◂.

Тема 1. ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

Проблема изучения взаимосвязей в экономике является одной из наиболее важнейших. При анализе таких взаимосвязей широкое распространение получило использование моделирования и количественного анализа. На базе последних выделилось и сформировалось направление исследований, получившее название эконометрика.

Термин «эконометрика» был введен в 1926 г. норвежским экономистом и статистиком Рагнаром Фришем. В буквальном переводе он означает «измерения в экономике». Однако область исследований данной дисциплины гораздо шире и глубже. Существуют различные варианты определения эконометрики – от чрезмерно широких до узко ориентированных. Ныне устоялся и широко распространен взгляд на эконометрику, который отражен в следующем определении.

Эконометрика – это наука, которая на основе статистических данных дает количественное выражение взаимосвязям экономических явлений и процессов.

Эконометрика, как научная дисциплина, зародилась и получила развитие на основе взаимодействия и объединения экономической теории, экономической статистики и математико-статистических методов.

Действительно, предмет исследования эконометрики, как и экономической теории, – это явления и процессы в экономике. Но в отличие от экономической теории эконометрика делает упор на количественные, а не на качественные аспекты этих явлений и процессов. Например, экономическая теория утверждает, что снижение цены товара приводит к увеличению спроса на данный товар (при неизменности всех прочих факторов). Но при этом практически неисследованным остается вопрос, как быстро и по какому закону происходит это возрастание. Эконометрика отвечает на данный вопрос для каждого конкретного случая.

Одной из основных задач экономической статистики является сбор, обработка и представление экономических данных. Эконометрика также активно пользуется этим инструментарием, но идет дальше, применяя его для анализа экономических взаимосвязей и прогнозирования.

Большинство эконометрических методов и приемов заимствовано из математической статистики. Однако методы математической статистики универсальны и не учитывают специфики экономических данных. Например, в экономике невозможно проводить многократные или управляемые эксперименты; многие показатели в экономике не могут принимать отрицательные или нулевые значения и т.д. Этот факт порождает ряд специфических проблем, решение которых не входит в задачи математической статистики.

1.1. ПОНЯТИЕ ЭКОНОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ

Первая же принципиальная идея, с которой встречается каждый изучающий экономику, – это идея о взаимосвязях между экономическими переменными (показателями, факторами и т.п.). Формирующийся на рынке спрос на некоторый товар рассматривается как функция его цены; затраты, связанные с изготовлением какого-либо продукта, предполагаются зависящими от объема производства; потребительские расходы могут быть функцией дохода и т.д. Все это – примеры связей между двумя переменными, одна из которых называется *объясняющей* (*независимой*, *экзогенной*) переменной, а другая – *объясняемой* (*результирующей*, *зависимой*, *эндогенной*) переменной. В приведенных примерах цена товара, объем производства, доход – объясняющие, или независимые, переменные; спрос на товар, затраты на производство товара, потребительские расходы – объясняемые, или зависимые, переменные. Приведенные примеры зависимостей между двумя экономическими показателями являются достаточно упрощенными.

Для большей реалистичности исследования характера изменения некоторого объясняемого (результирующего) показателя следует рассматривать его зависимость от нескольких объясняющих переменных. Например, спрос на товар можно рассматривать как функцию его цены, потребительского дохода и цен на конкурирующие и дополняющие товары; производственные затраты будут зависеть от объема производства и цен на основные производственные ресурсы.

Количественное изучение экономических явлений и процессов осуществляется с помощью *математических моделей* – абстракций реальных объектов (явлений, процессов), в которых интересующие исследователя отношения между реальными элементами заменены подходящими отношениями между математическими категориями. Эти отношения, как правило, представлены в виде уравнений и (или) неравенств между переменными, характеризующими функционирование реального моделируемого объекта. Количество отношений (связей), включаемых в модель, зависит, как правило, от условий, при которых эта модель конструируется, и от подробности объяснений, к которым мы стремимся.

Если речь идет о математической модели, описывающей механизм функционирования некоторой экономической или социально-экономической системы, то такую модель принято называть *экономико-математической* или просто *экономической*.

Не всякая экономико-математическая модель может считаться эконометрической. Данная особенность отражена в следующем определении. *Эконометрическая модель* – это такая экономико-математическая модель, что:

• она имитирует механизм функционирования гипотетического, а не конкретного явления (или системы), имеющего случайную природу;

• значения отдельных ее характеристик (параметров) оцениваются по результатам наблюдений, характеризующим функционирование конкретного, а не гипотетического моделируемого явления (или системы).

Процесс построения эконометрической модели называется *эконометрическим моделированием*.

1.2. ОСНОВНЫЕ ЭТАПЫ ЭКОНОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ

Можно выделить шесть основных этапов эконометрического моделирования: постановочный, априорный, этап параметризации, информационный, этап идентификации и верификации модели.

1-й этап (*постановочный*) – определение конечных целей моделирования, набора участвующих в модели факторов и показателей (переменных), их роли.

2-й этап (*априорный*, *предмодельный*) – анализ экономической сущности изучаемого явления, формирование и формализация априорной информации об этом явлении в виде ряда гипотез и исходных допущений;

3-й этап (*параметризация*) – собственно моделирование, т.е. выбор общего вида модели, в том числе состава и формы входящих в нее связей;

4-й этап (*информационный*) – сбор необходимой статистической информации, т.е. регистрация значений участвующих в модели факторов и показателей на различных временных или пространственных тактах функционирования изучаемого явления;

5-*й этап* (*идентификация модели*) – статистический анализ модели и статистическое оценивание ее неизвестных параметров;

6-й этап (*верификация модели*) – сопоставление реальных и модельных данных для проверки истинности, адекватности модели.

Математическая модель, в том числе математическая модель экономического явления или процесса, может быть сформулирована на общем (качественном) уровне, без «настройки» на конкретные статистические данные, т.е. она может иметь смысл и без 4-го и 5-го этапов. Только тогда она не является эконометрической. Суть именно эконометрической модели заключается в том, что она, будучи представленной в виде набора математических соотношений, описывает функционирование конкретной экономической системы, а не системы вообще. Поэтому ее параметры оцениваются с помощью конкретных статистических данных, что предусматривает обязательную реализацию 4-го и 5-го этапов моделирования.

Контрольные вопросы

1. Дайте определение эконометрики.
2. На базе каких дисциплин сформировалась и развилась эконометрика? Какой вклад вносят эти дисциплины в эконометрику?
3. Что такое экзогенные и эндогенные переменные, как их можно назвать иначе?
4. Что такое математическая модель? Экономико-математическая модель?
5. В чем состоит отличие эконометрической модели от экономико-математической?
6. Перечислите основные этапы эконометрического моделирования.

Литература

1. *Айвазян С.А.*, *Мхитарян В.С.* Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.
2. *Кремер Н.Ш.*, *Путко Б.А.* Эконометрика. – М.: ЮНИТИ-ДАНА, 2002. – Гл. 1.
3. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – Гл. 1.

Тема 2. ЛИНЕЙНАЯ ЭКОНОМЕТРИЧЕСКАЯ МОДЕЛЬ С ДВУМЯ ПЕРЕМЕННЫМИ

2.1. ВЗАИМОСВЯЗИ ЭКОНОМИЧЕСКИХ ПЕРЕМЕННЫХ

Проблема изучения взаимосвязей экономических переменных является одной из наиболее важных в экономическом анализе. Любая экономическая политика заключается в регулировании экономических показателей, и она должна основываться на знании того, как эти показатели влияют на другие показатели, являющиеся ключевыми для принимающего решение политика или предпринимателя. Так, в рыночной экономике нельзя непосредственно регулировать темп инфляции, но на него можно воздействовать средствами бюджетной, налоговой и кредитно-денежной политики.

В естественных науках большей частью имеют дело со строгими (функциональными) зависимостями, при которых каждому значению одной переменной соответствует единственное значение другой. Однако в подавляющем большинстве случаев между экономическими переменными таких закономерностей нет. Например, нет строгой зависимости между доходом и потреблением, ценой и спросом, производительностью труда и стажем работы и т.д. Это связано с целым рядом причин:

- при анализе влияния одной переменной на другую не учитываются многие другие факторы, воздействующие на нее;
- влияние одной переменной на другую может быть не прямым, а проявляться через цепочку других факторов;
- воздействие одних переменных на другие носят случайный характер и т.д.

Зависимость переменных при наличии воздействия случайных факторов, при которой изменение одной из величин влечет изменение закона распределения другой, называется *статистической* (или *стохастической*, *вероятностной*).

В наиболее общем виде при изучении статистических взаимосвязей исследователя интересует количественная оценка их наличия и направления, а также характеристика силы и формы влияния одних факторов на другие. Для решения этих задач применяются две группы методов, одна из которых включает в себя методы корреляционного анализа, а другая – регрессионного анализа. Часто эти методы объединяют в корреляционно-регрессионный анализ, что объясняется наличием целого ряда общих вычислительных процедур, взаимодополнением при интерпретации результатов и др.

Можно указать два варианта рассмотрения статистических взаимосвязей между двумя переменными X и Y. В первом случае обе переменные считаются равноправными в том смысле, что они не подразделяются на первичную и вторичную (независимую и зависимую) переменные. Такую статистическую зависимость называют *корреляционной*. Основным в этом случае является вопрос о наличии и силе взаимосвязи между рассматриваемыми переменными, что является задачей *корреляционного анализа*, основной мерой которого является коэффициент корреляции.

Другой вариант рассмотрения взаимосвязей выделяет одну из величин как независимую (объясняющую), а другую как зависимую (объясняемую). В этом случае изменение первой из них может служить причиной для изменения другой. Например, рост дохода ведет к увеличению потребления; рост цены – к снижению спроса; снижение процентной ставки увеличивает инвестиции и т.д. Однако такая зависимость не является однозначной в том смысле, что каждому конкретному значению объясняющей переменной может соответствовать не одно, а много значений зависимой переменной из некоторой области. Другими словами, каждому конкретному значению объясняющей переменной X соответствует некоторое вероятностное распределение зависимой переменной Y, рассматриваемой как случайная величина. Поэтому анализируют влияние объясняющей переменной X на зависимую переменную Y в «среднем», т.е. изучают закономерность в изменении условного математического ожидания $M(Y\ /\ X = x)$ – математического ожидания случайной переменной Y, вычисленного в предположении, что переменная X приняла значение x, – как функции от x. Зависимость такого типа, выражаемая соотношением

$$M(Y / X = X) = F(X), \qquad (2.1)$$

называется регрессионной зависимостью, или регрессией Y на X. При этом переменная X называется независимой, объясняющей, экзогенной переменной, или регрессором, а Y – зависимой, объясняемой, эндогенной переменной, или результирующим признаком. Уравнение (2.1) называется также уравнением регрессии переменной Y на X , функция f(x) – функцией регрессии Y на X, а ее график – кривой регрессии.

При рассмотрении регрессионной зависимости двух переменных говорят о парной регрессии, а нескольких – о множественной регрессии. Основными задачами регрессионного анализа (парного и множественного) являются установление вида регрессионной зависимости между рассматриваемыми переменными, оценка функции регрессии и прогноз значений зависимой переменной.

Природа регрессионной зависимости может носить двойственный характер:

• регистрация результирующего показателя Y неизбежно связана с некоторыми случайными ошибками измерения ε, в то время как объясняющая переменная X измеряется без ошибок или величины этих ошибок пренебрежимо малы по сравнению с соответствующими ошибками измерения результирующего показателя;

• значения результирующего показателя Y зависят не только от соответствующих значений объясняющей переменной, но еще и от ряда неконтролируемых факторов; поэтому при каждом фиксированном значении объясняющей переменной соответствующие значения результирующего показателя Y неизбежно подвержены некоторому случайному разбросу ε.

Удобной формой записи такого рода зависимости между результирующей и объясняющей переменными является соотношение вида:

$$Y = F(X) + \varepsilon, \tag{2.2}$$

называемое регрессионной моделью.

Случайная переменная ε в модели (2.2) отражает случайную природу Y и характеризует отклонение зависимой переменной Y от функции регрессии f(X). Эту переменную называют возмущением (либо ошибкой, отклонением).

2.2. ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ И МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Предположим, что функция регрессии переменной Y на X линейна:

$$y = M(Y / X = x) = a\,x + b. \tag{2.3}$$

В этом случае говорят о *линейной регрессии*. Тогда *модель парной линейной регрессии* можно представить в виде

$$Y = aX + b + \varepsilon, \tag{2.4}$$

где a, b – теоретические параметры (теоретические коэффициенты) регрессии;

ε – случайное возмущение.

Модель линейной регрессии является наиболее распространенным (и простым) видом зависимости между экономическими переменными. Построенная линейная модель может служить начальной точкой эконометрического анализа. Кроме того, линейная регрессия находит широкое применение в эконометрике в силу четкой экономической интерпретации ее параметров.

Знак коэффициента регрессии a в модели указывает направление связи между переменными X и Y: если a > 0, связь прямая; если a < 0, то связь обратная. Величина коэффициента регрессии a показывает, на какую величину в среднем изменится результирующая переменная Y, если переменная X увеличится на одну единицу своего измерения.

Формально значение параметра b в модели – это среднее значение результирующей переменной Y при X = 0. Однако здесь необходима определенная осторожность. Если переменная X не имеет и не может иметь нулевого значения, вышеуказанная трактовка параметра b теряет смысл.

Пример 2.1. Парная линейная регрессия часто используется в эконометрике при изучении зависимости потребления Y от величины дохода X.

Предположим, что функция регрессии потребления на доход (в тыс. руб.) имеет вид: $y^* = 1{,}5 + + 0{,}65x$.

Коэффициент регрессии a = 0,65 характеризует склонность к потреблению. Он показывает, что из каждой тысячи рублей дохода на потребление расходуется в среднем 650 рублей.

Свободный член b = 1,5 формально показывает, что при отсутствии дохода (x = 0) потребление, тем не менее, составляет 1 500 рублей. Если данный вывод применим к отдельному индивидууму или отдельной семье (они могут жить и в долг), то он вряд ли применим, например, к экономике государства. ■

Для точного описания уравнения регрессии (2.3) необходимо знать условный закон распределения зависимой переменной Y при условии, что независимая переменная X примет значение x, т.е. $X = x$. В статистической практике такую информацию получить, как правило, не удается. Однако у исследователя имеются *статистические* (или *выборочные*) *данные* (x_i, y_i), $i = 1, 2, \dots, n$, полученные в результате n независимых наблюдений над переменными X и Y. В этом случае по выборочным данным можно *оценить* (*приближенно выразить*) теоретическую функцию регрессии (2.3). Такой оценкой является *выборочное уравнение регрессии*

$$y^* = a^* \cdot x + b^*, \tag{2.5}$$

где y^* – оценка условного математического ожидания переменной Y, т.е. функции $y = M(Y / X = x)$;

a^* и b^* – оценки неизвестных параметров a и b соответственно.

Построение выборочного уравнения регрессии (2.5) сводится к нахождению его параметров a^* и b^*, которые могут быть найдены разными способами. Наиболее простой метод состоит в следующем. Изобразим выборочные данные (x_i, y_i), $i = 1, 2, \dots, n$, в виде точек на координатной плоскости (рис. 2.1). Такое графическое изображение статистических данных называется *корреляционным полем* (или *диаграммой рассеяния*).

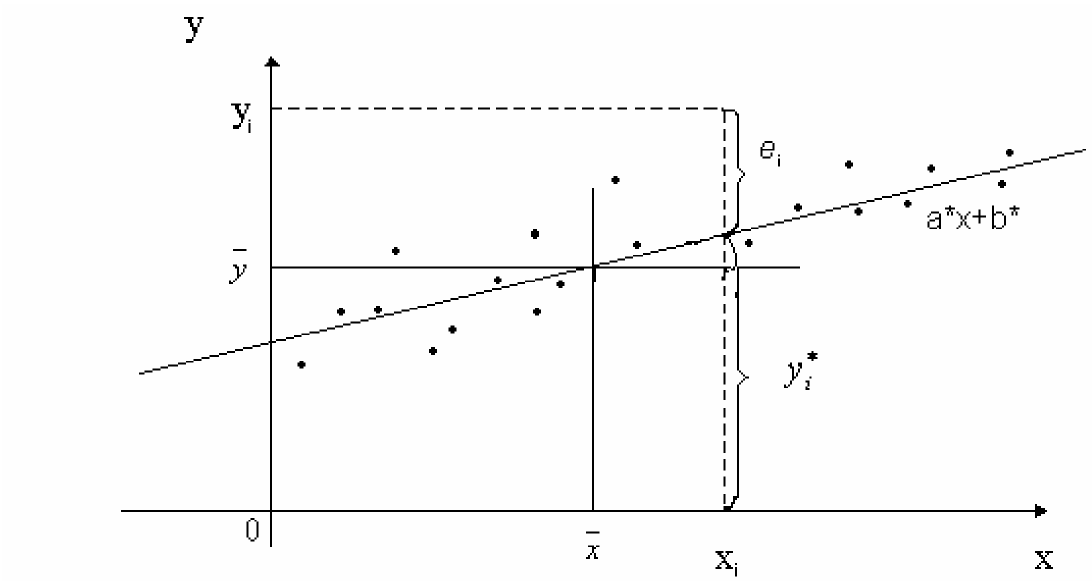


Рис. 2.1. Корреляционное поле

Выберем две точки на диаграмме рассеяния и проведем через них прямую линию. Далее по графику нетрудно определить коэффициенты этой прямой. Очевидно, подобным образом можно построить много различных прямых, однако непонятно, какая из них наилучшим образом отражает зависимость между переменными X и Y.

Самый распространенный и теоретически обоснованный подход к оцениванию параметров линейной регрессии основан на *методе наименьших квадратов* (*МНК*). Этот метод является также наиболее простым с вычислительной точки зрения. Кроме того, оценки, полученные МНК, при определенных предпосылках, обладают рядом оптимальных свойств.

Суть МНК нахождения неизвестных параметров a^* и b^* состоит в том, чтобы сумма квадратов отклонений выборочных значений y_i от значений y_i^*, найденных по уравнению регрессии (2.5), *была минимальной:*

$$G = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - y_i^*\right)^2 = \sum_{i=1}^{n} \left(y_i - a^* x_i - b^*\right)^2 \to \min. \tag{2.6}$$

Необходимым условием существования минимума функции двух переменных $G = G(a^*, b^*)$ является равенство нулю ее частных производных:

$$\begin{cases} \dfrac{\partial G}{\partial a^*} = -2\sum\limits_{i=1}^{n}\left(y_i - a^* x_i - b^*\right) = 0, \\ \dfrac{\partial G}{\partial b^*} = -2\sum\limits_{i=1}^{n}\left(y_i - a^* x_i - b^*\right) x_i = 0. \end{cases}$$

В уравнениях системы раскроем скобки и после несложных преобразований получим так называемую *систему нормальных уравнений* для определения неизвестных параметров линейной регрессии:

$$\begin{cases} b^* \cdot n + a^* \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i, \\ b^* \sum\limits_{i=1}^{n} x_i + a^* \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases}$$

Решение системы можно найти достаточно легко методом последовательного исключения неизвестных (или методом Крамера):

$$a^* = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}, \tag{2.7}$$

$$b^* = \overline{y} - a^* \cdot \overline{x}. \tag{2.8}$$

В соотношениях (2.7) и (2.8) соответствующие средние определяются по формулам:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \tag{2.9}$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i; \tag{2.10}$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2; \tag{2.11}$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i. \tag{2.12}$$

Подставляя значение b^* из равенства (2.8) в уравнение регрессии (2.5), получим

$$y^* = \overline{y} - a^* \cdot \overline{x} + a^* \cdot x$$

или

$$y^* - \overline{y} = a^*(x - \overline{x}). \tag{2.13}$$

Из полученного уравнения регрессии (2.13) видно, что линия регрессии проходит через точку $(\overline{x}, \overline{y})$ (см. рис. 2.1).

Пример 2.2. Изучается зависимость между величиной потребительских расходов Y (усл. ед.) и доходом X (усл. ед.) на одного члена семьи. Для этого помесячно в течение года отобрана выборка объема n = 12, результаты которой приведены в следующей таблице.

Т а б л и ц а 2.1

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150
y_i	102	105	108	110	115	117	119	125	132	130	141	144

Необходимо:

а) определить вид зависимости между X и Y и записать ее в виде модели;

б) по МНК оценить параметры уравнения регрессии Y на X и построить выборочное уравнение регрессии;

в) дать экономическую интерпретацию полученным оценкам параметров уравнения регрессии.

Решение. а) Для определения вида зависимости между X и Y построим корреляционное поле (рис. 2.2).

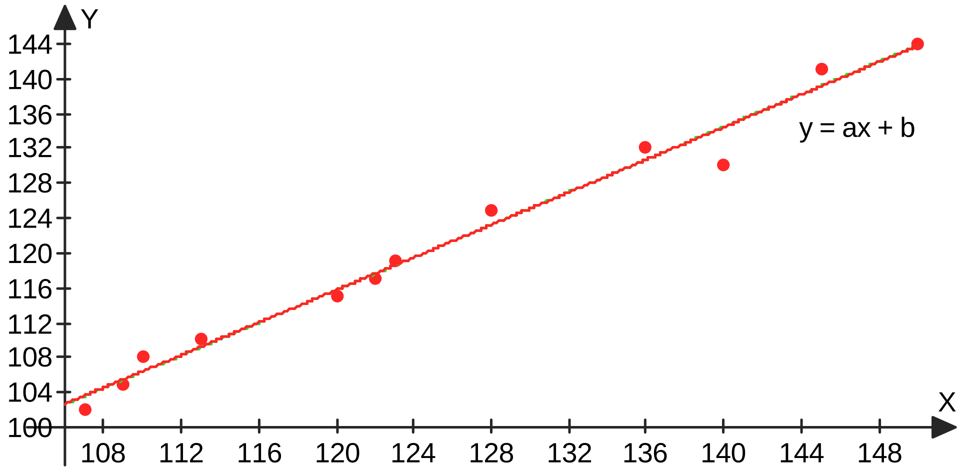


Рис. 2.2. Корреляционное поле и прямая регрессии

Поскольку точки на корреляционном поле располагаются около некоторой прямой $y = ax + b$, то можно предположить наличие линейной регрессионной зависимости между переменными X и Y:

$$Y = aX + b + \varepsilon,$$

где a, b – неизвестные параметры; ε – случайное возмущение.

б) Для наглядности вычислений построим табл. 2.2.

Таблица 2.2

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	107	102	11 449	10 404	10 914
2	109	105	11 881	11 025	11 445
3	110	108	12 100	11 664	11 880
4	113	110	12 769	12 100	12 430
5	120	115	14 400	13 225	13 800
6	122	117	14 884	13 689	14 274
7	123	119	15 129	14 161	14 637
8	128	125	16 384	15 625	16 000
9	136	132	18 496	17 424	17 952
10	140	130	19 600	16 900	18 200
11	145	141	21 025	19 881	20 445
12	150	144	22 500	20 736	21 600
Σ	1 503	1 448	190 617	176 834	183 577

Вычислим по формулам (2.9) – (2.12) все необходимые средние:

$$\overline{x} = \frac{1}{12}\sum_{i=1}^{12} x_i = \frac{1503}{12} = 125{,}25; \quad \overline{y} = \frac{1}{12}\sum_{i=1}^{12} y_i = \frac{1448}{12} = 120{,}667;$$

$$\overline{x^2}=\frac{1}{12}\sum_{i=1}^{12}x_i^2=\frac{190617}{12}=15884{,}75;\ \ \overline{x\cdot y}=\frac{1}{12}\sum_{i=1}^{12}x_iy_i=\frac{183577}{12}=15298{,}083.$$

Согласно МНК по формулам (2.8) и (2.9) находим

$$a^*=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{x^2}-\overline{x}^2}=\frac{15298{,}08-125{,}25\cdot 120{,}67}{15884{,}75-(125{,}25)^2}=\frac{184{,}1625}{197{,}1875}=0{,}9339;$$
$$b^*=\overline{y}-a^*\cdot\overline{x}=120{,}67-0{,}9339\cdot 125{,}25=3{,}699.$$

Таким образом, уравнение парной линейной регрессии имеет вид:

$$y^* = 0{,}9339x + 3{,}699. \tag{2.14}$$

Изобразим данную прямую на корреляционном поле (рис. 2.2).

в) коэффициент $a^* \approx 0{,}934$ полученного уравнения регрессии показывает, что увеличение дохода (на одного человека) на 1 усл. ед. влечет увеличение расходов примерно на 0,934 усл. ед. Значение $b^* \approx 3{,}7$ говорит о том, что при нулевом доходе расходы составят в среднем примерно 3,7 усл. ед. Этот факт можно объяснить для отдельной семьи (она может тратить накопленные или одолженные средства), но для совокупности семей он теряет смысл. ■

2.3. ПРЕДПОСЫЛКИ ЛИНЕЙНОГО РЕГРЕССИОННОГО АНАЛИЗА. СВОЙСТВА МНК-ОЦЕНОК

Метод наименьших квадратов позволяет определить оценки параметров регрессии и с их помощью построить уравнение регрессии. Однако, являясь лишь приближенными значениями, они не позволяют сделать вывод, насколько близки оценки a^* и b^* к своим теоретическим прототипам a и b, насколько точно выборочное уравнение регрессии соответствует теоретическому, насколько надежны найденные оценки. Для ответа на эти вопросы и получения наилучших МНК-оценок необходимо, чтобы выполнялся ряд предпосылок относительно случайного возмущения ε.

Предположим, что над переменными X и Y, связанными линейной регрессионной зависимостью (2.4), проведено n наблюдений, результатом которых является n пар значений переменных (x_i, y_i), $i = 1, 2, \ldots, n$. Тогда для каждого наблюдения будет иметь место зависимость вида

$$y_i = a\cdot x_i + b + \varepsilon_i, \tag{2.15}$$

где ε_i – случайное возмущение i -го наблюдения, $i = 1, 2, \ldots, n$.

Основные предпосылки регрессионного анализа

1^0. В модели (2.4) парной линейной регрессии возмущение ε есть случайная величина, объясняющая переменная X – величина неслучайная.

2^0. Математическое ожидание случайного возмущения ε_i равно нулю:

$M(\varepsilon_i) = 0$ для всех [1)].

Данное условие означает, что хотя в каждом конкретном наблюдении случайное отклонение может быть положительным или отрицательным, оно не должно иметь систематического смещения. Другими словами, случайное возмущение в среднем не должно оказывать влияния на результирующую переменную.

3^0. Дисперсия случайных возмущений постоянна:

$D(\varepsilon_i) = \sigma^2$ для всех наблюдений.

Выполнение данного условия называется *гомоскедастичностью* (*постоянством дисперсии*) возмущений, невыполнение – *гетероскедастичностью* (*непостоянством дисперсии*) возмущений.

4^0. Возмущения ε_i и ε_j $(i \neq j)$ не коррелированны:

$cov(\varepsilon_i, \varepsilon_j) = 0$ для всех наблюдений.

Если данное условие выполняется, то говорят об отсутствии *автокорреляции* возмущений, в противном случае – о ее наличии.

---

1) Во всех предпосылках и в дальнейшем $i = 1, 2, \ldots, n$.

5^0. Возмущение ε_i есть нормально распределенная случайная величина:
$\varepsilon_i \sim N(0, \sigma)$ *для всех наблюдений*.

Замечание 2.1. При выполнении предпосылки 5^0 условие 4^0 некоррелированности возмущений ε_i и ε_j $(i \neq j)$ равносильно их независимости. ◂

***Теорема 2.1* (*Гаусса – Маркова*).** *Если выполнены предпосылки $1^0 – 4^0$, то МНК-оценки a^* и b^* обладают следующими свойствами*:

- *оценки являются несмещенными*, *т.е.* $M(a^*) = a$, $M(b^*) = b$;

$$D(a^*) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}, \quad D(b^*) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \cdot \frac{\sum x_i^2}{n}, \tag{2.16}$$

где σ^2 – дисперсия возмущений;

- *оценки состоятельны: $a^* \to a$, $b^* \to b$ по вероятности при $n \to \infty$;*
- *оценки эффективны*, *т.е. они имеют наименьшие дисперсии по сравнению с любыми другими несмещенными оценками*, *линейными относительно величин y_i*, $i = 1, 2, \ldots, n$.

Оценкой теоретического уравнения регрессии (2.3) является выборочное уравнение регрессии (2.5), где a^* и b^* – МНК-оценки, найденные по формулам (2.7) и (2.8) соответственно. Воздействие случайных возмущений ε_i в модели (2.15) определяется с помощью дисперсии возмущений или *остаточной дисперсии* $\sigma^2 = D(\varepsilon_i)$. Так как случайные возмущения ε_i по выборке определены быть не могут, они заменяются на отклонения $e_i = y_i - a^* x_i - b^*$ фактических значений y_i переменной Y от оцененных с помощью уравнения регрессии (2.5). Тогда несмещенной оценкой дисперсии возмущений (остаточной дисперсии) σ^2 является *выборочная остаточная дисперсия*

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left(y_i - y_i^*\right)^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2, \tag{2.17}$$

где $y_i^* = a^* \cdot x_i + b^*$ – значение зависимой переменной Y, найденное по выборочному уравнению регрессии (2.5);

$e_i = y_i - y_i^*$ – выборочная оценка возмущения ε_i или остаток регрессии.

Оценками дисперсий выборочных коэффициентов регрессии a^* *и* b^*, в силу (2.16) и (2.17), являются величины

$$S_{a^*}^2 = \frac{S^2}{\sum (x_i - \bar{x})^2}, \tag{2.18}$$

$$S_{b^*}^2 = \frac{S^2}{\sum (x_i - \bar{x})^2} \cdot \frac{\sum x_i^2}{n}. \tag{2.19}$$

Выборочные стандартные отклонения $S_{a^*} = \sqrt{S_{a^*}^2}$ и $S_{b^*} = \sqrt{S_{b^*}^2}$ выборочных коэффициентов регрессии a^* и b^* соответственно часто называют *стандартными ошибками коэффициентов регрессии*.

Пример 2.3. По данным примера 2.2 вычислить выборочную остаточную дисперсию и стандартные ошибки коэффициентов регрессии.

Решение. Для вычисления выборочной остаточной дисперсии S^2 составим следующую таблицу:

Т а б л и ц а 2.3

i	x_i	y_i	$y_i^* = 0{,}9339x_i + 3{,}699$	$e_i = y_i - y_i^*$	e_i^2
1	2	3	4	5	6
1	107	102	103,63	–1,63	2,66
2	109	105	105,49	–0,49	0,24
3	110	108	106,43	1,57	2,46
4	113	110	109,23	0,77	0,59
5	120	115	115,77	–0,77	0,59
6	122	117	117,63	–0,63	0,40
7	123	119	118,57	0,43	0,18
8	128	125	123,24	1,76	3,10
9	136	132	130,71	1,29	1,66
10	140	130	134,45	–4,45	19,8
11	145	141	139,11	1,89	3,57
12	150	144	143,78	0,22	0,05
Σ	1503	1448	≈ 1448*)	≈ 0*)	35,3

По формуле (2.17) имеем теперь:

$$S^2 = \frac{35{,}3}{12-2} = 3{,}53\text{, откуда } S = \sqrt{3{,}53} \approx 1{,}8788\,.$$

В примере 2.2 найдены величины $\sum_{i=1}^{12} x_i^2 = 190617$ и $\overline{x} = 125{,}25$.

При вычислении $\sum_{i=1}^{12}\left(x_i - \overline{x}\right)^2$ воспользуемся табл. 2.2. Для этого заметим, что

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}(\overline{x})^2 = \sum_{i=1}^{n} x_i^2 - 2\overline{x}\cdot n\cdot\frac{1}{n}\sum_{i=1}^{n} x_i + n\cdot(\overline{x})^2 =$$
$$= \sum_{i=1}^{n} x_i^2 - (\overline{x})^2,$$

поэтому

$$\sum_{i=1}^{12}\left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{12} x_i^2 - 12\cdot(\overline{x})^2 = 190\,617 - 12\cdot (125{,}25)^2 = 2\,366{,}256.$$

По формулам (2.18) и (2.19) находим

$$S_{a^*}^2 = \frac{3{,}53}{2366{,}2566} = 0{,}0015;\quad S_{b^*}^2 = \frac{3{,}53}{2366{,}2566}\cdot\frac{190617}{12} = 23{,}697\,,$$

откуда

$$S_{a^*} = \sqrt{0{,}0015} = 0{,}0387;\quad S_{b^*} = \sqrt{23{,}697} = 4{,}868\,.\ \blacksquare$$

*) Значения округляются до сотых и учитываются погрешности округлений.

2.4. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ В ПАРНОЙ РЕГРЕССИИ

В последующих формулах для упрощения записи знак суммы $\sum_{i=1}^{n}$ будем писать без индексов: $\sum$, предполагая, что суммирование ведется по i от 1 до n. Также для переменных с индексом i будем подразумевать, что $i = 1, 2, \ldots, n$ (если не указано иное).

2.4.1. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ ПАРАМЕТРОВ МОДЕЛИ

Построим доверительные интервалы для параметров a, b и σ^2 линейной регрессионной модели (2.4), которые с заданной надежностью (доверительной вероятностью) $\gamma = 1 - \alpha$ накрывают неизвестные значения a, b и σ^2. Для этого нам понадобится следующее утверждение.

Теорема 2.2. *Если выполнены предпосылки $1^0 - 5^0$, то*:

- *статистика*

$$t_{a^*} = \frac{a^* - M(a^*)}{S_{a^*}} = \frac{(a^* - a)}{S}\sqrt{\sum(x_i - \overline{x})^2}$$

имеет t-распределение Стьюдента с $(n-2)$ *степенями свободы*;

- *статистика*

$$t_{b^*} = \frac{b^* - M(b^*)}{S_{b^*}} = \frac{(b^* - b)}{S}\sqrt{\frac{n\sum(x_i - \overline{x})^2}{\sum x_i^2}}$$

имеет t-распределение Стьюдента с $(n-2)$ *степенями свободы*;

- *статистика*

$$\chi^2 = \frac{\sum e_i^2}{\sigma^2} = \frac{(n-2)S^2}{\sigma^2}$$

имеет распределение хи-квадрат (χ^2*-распределение*) *с* $(n-2)$ *степенями свободы*.

Здесь $S = \sqrt{S^2}$ *– выборочное остаточное среднеквадратическое* (*стандартное*) *отклонение*.

Пусть $t_{1-\alpha/2}(n-2)$ – квантиль порядка $(1 - \alpha/2)$ распределения Стьюдента с $(n-2)$ степенями свободы;

$\chi^2_{\alpha/2}(n-2)$ и $\chi^2_{1-\alpha/2}(n-2)$ – квантили порядков $\alpha/2$ и $(1 - \alpha/2)$ соответственно χ^2-распределения с $(n-2)$ степенями свободы.

Значения этих величин при известных n и $\alpha = 1 - \gamma$ можно найти в Приложении по таблицам 1 и 2 соответственно.

Используя стандартную процедуру построения доверительных интервалов, когда известно распределение соответствующей статистики, получим следующие *интервальные оценки параметров* a, b и σ^2 надежности $\gamma = 1 - \alpha$:

$$a^* - t_{1-\alpha/2}(n-2)\cdot S_{a^*} < a < a^* + t_{1-\alpha/2}(n-2)\cdot S_{a^*}; \tag{2.20}$$

$$b^* - t_{1-\alpha/2}(n-2)\cdot S_{b^*} < b < b^* + t_{1-\alpha/2}(n-2)\cdot S_{b^*}; \tag{2.21}$$

$$\frac{nS^2}{\chi^2_{1-\alpha/2}(n-2)} < \sigma^2 < \frac{nS^2}{\chi^2_{\alpha/2}(n-2)}. \tag{2.22}$$

2.4.2. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ФУНКЦИИ РЕГРЕССИИ

Построение доверительного интервала для функции регрессии, т.е. для условного математического ожидания $y = M(Y/X = x)$, основывается на следующей теореме.

Теорема 2.3. *Если выполнены предпосылки $1^0 - 5^0$, то статистика*

$$t_{y^*} = \frac{y^* - M(y^*)}{S_{y^*}} = \frac{y^* - y}{S\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum (x_i - \bar{x})^2}}},$$

где $y^* = a^* \cdot x + b^*$ – выборочное уравнение регрессии, при любом $x \in R$ имеет t-распределение Стьюдента с $(n - 2)$ степенями свободы.

Воспользовавшись снова стандартной процедурой построения доверительного интервала, когда известно распределение соответствующей статистики, получим следующую *интервальную оценку уравнения регрессии*

$y = a\,x + b$ надежности $\gamma = 1 - \alpha$:

$$y^* - t_{1-\alpha/2}(n-2) \cdot S_{y^*} < y < y^* + t_{1-\alpha/2}(n-2) \cdot S_{y^*}. \qquad (2.23)$$

Здесь $y^* = a^* \cdot x + b^*$ – выборочное уравнение регрессии;

$S_{y^*} = S\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum (x_i - \bar{x})^2}}$ – выборочное среднеквадратическое отклонение (стандартная ошибка) уравнения регрессии;

S – выборочное остаточное среднеквадратическое отклонение.

Из формулы (2.23) видно, что величина (длина) доверительного интервала зависит от значения x объясняющей переменной: при $x = \bar{x}$ она минимальна, а по мере удаления x от $\bar{x}$ величина доверительного интервала увеличивается (рис. 2.3). Множество доверительных интервалов, содержащихся между *нижней* и *верхней доверительными границами* неравенства (2.23), называется *доверительной полосой* для теоретического уравнения регрессии (на рис. 2.3 – это множество, содержащееся между сплошными кривыми).

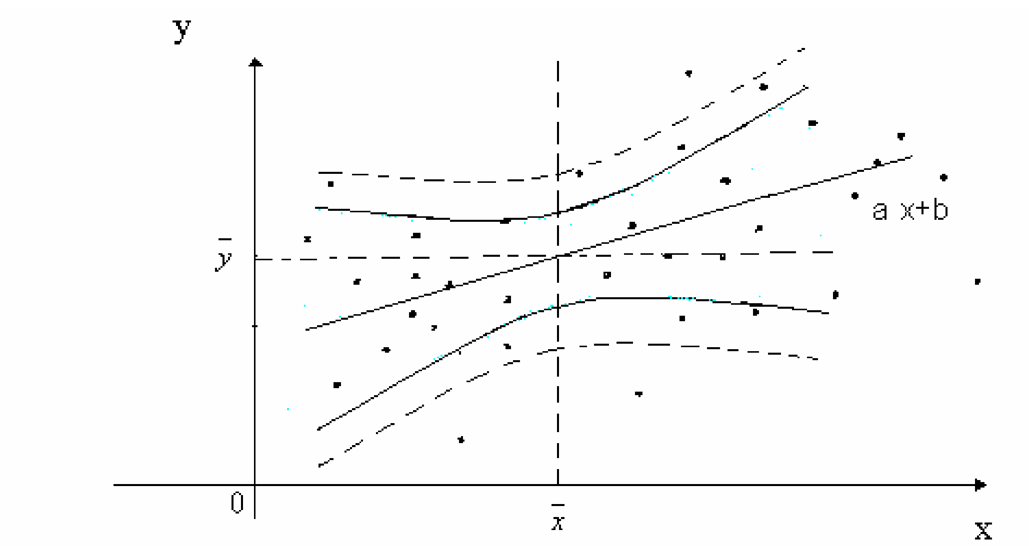


Рис. 2.3. Доверительные полосы

2.4.3. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ИНДИВИДУАЛЬНЫХ ЗНАЧЕНИЙ ЗАВИСИМОЙ ПЕРЕМЕННОЙ

Построенная доверительная полоса для уравнения регрессии $y = a\,x + b$ (рис. 2.3) определяет местоположение теоретической линии регрессии, но не отдельных возможных значений зависимой переменной $Y_x = a\,x + b + \varepsilon$, которые могут отклоняться от линии регрессии. Определение доверительного интервала для этих величин основывается на другой теореме.

Теорема 2.4. *Пусть* $Y_x^* = a^* x + b^* + \varepsilon$ *. Если выполнены предпосылки* $1^0 - 5^0$*, то статистика*

$$t_{Y_x^*} = \frac{Y_x^* - M\left(Y_x^*\right)}{S_{Y_x^*}} = \frac{Y_x - y^*}{S\sqrt{1 + \frac{1}{n} + \frac{\left(x - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}},$$

где $y^* = a^* \cdot x + b^*$ – выборочное уравнение регрессии, при любом x $\in$R имеет t-распределение Стьюдента с (n – 2) степенями свободы.

Доверительный интервал для отдельных возможных значений зависимой переменной Y строится аналогично п. 2.4.2 и определяется по формуле:

$$y^* - t_{1-\alpha/2}(n-2)\cdot S_{Y_x^*} < Y_x < y^* + t_{1-\alpha/2}(n-2)\cdot S_{Y_x^*}. \qquad (2.24)$$

Здесь $y^* = a^* \cdot x + b^*$ – выборочное уравнение регрессии;

$S_{Y_x^*} = S\sqrt{1 + \frac{1}{n} + \frac{\left(x - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}$ – выборочное среднеквадратическое

отклонение (стандартная ошибка) индивидуальных значений зависимой переменной;

S – выборочное остаточное среднеквадратическое отклонение.

Как видно из (2.24) доверительная полоса для отдельных значений зависимой переменной Y шире, чем для уравнения регрессии (на рис. 2.3 – это множество, содержащееся между пунктирными кривыми).

2.5. ПРОВЕРКА КАЧЕСТВА РЕГРЕССИОННОЙ МОДЕЛИ

2.5.1. ОЦЕНКА ЗНАЧИМОСТИ ПАРАМЕТРОВ МОДЕЛИ

С помощью метода наименьших квадратов мы получили лишь оценки параметров уравнения регрессии. Важной задачей на начальном этапе статистического анализа построенной модели является проверка значимости параметров регрессии. Чтобы проверить, *значимы* ли эти параметры (т.е. значительно ли они отличаются от нуля в «истинном» уравнении регрессии

$y = a\,x + b$), используют статистические методы проверки гипотез.

Предположим, что в качестве основной гипотезы H_0 выдвигается гипотеза о незначимом отличии от нуля «истинного» коэффициента регрессии a: $H_0 = \{a = 0\}$; альтернативной гипотезой H_1 при этом является гипотеза о значимости данного коэффициента: $H_1 = \{a \neq 0\}$. Для проверки гипотезы H_0 используется статистика

$$t_{a^*} = \frac{a^*}{S_{a^*}} = a^* \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{S^2}}, \qquad (2.25)$$

которая при справедливости H_0 имеет распределение Стьюдента с (n – 2) степенями свободы (теорема 2.2). Следовательно, нулевая гипотеза H_0: a = 0 отклоняется на основании t-критерия Стьюдента и принимается гипотеза о значимости коэффициента регрессии, если

$$\left|t_{a^*}\right| \geq t_{1-\alpha/2}(n-2),$$

где α – требуемый уровень значимости, $t_{1-\alpha/2}$ (n – 2) – квантиль порядка ($1-\alpha/2$) распределения Стьюдента с (n – 2) степенями свободы. При невыполнении данного неравенства считается, что нет оснований для отклонения H_0.

По аналогичной схеме проверяется гипотеза о значимости свободного члена b. Для этого используется статистика

$$t_{b^*} = \frac{b^*}{S_{b^*}} = b^* \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{S^2} \cdot \frac{n}{\sum x_i^2}}, \qquad (2.26)$$

которая при справедливости нулевой гипотезы H_0: $b = 0$ имеет распределение Стьюдента с $(n - 2)$ степенями свободы (теорема 2.2). Следовательно, H_0 отклоняется на основании t-критерия Стьюдента и принимается гипотеза о значимости свободного члена уравнения регрессии, если

$$\left| t_{b^*} \right| \geq t_{1-\alpha/2}(n-2),$$

где α – требуемый уровень значимости.

Замечание 2.2. При оценке значимости параметров линейной регрессии на начальном этапе можно использовать следующее грубое правило, позволяющее не прибегать к таблицам:

• если $\left| t_{a^*} \right| \leq 1$ ($\left| t_{b^*} \right| \leq 1$), то коэффициент регрессии a (свободный член b) не может быть признан значимым, так как соответствующая надежность составит менее 0,7;

• если $1 < \left| t_{a^*} \right| \leq 2$ ($1 < \left| t_{b^*} \right| \leq 2$), то параметр регрессии может рассматриваться как относительно (слабо) значимый: надежность в этом случае лежит между значениями 0,7 и 0,95;

• если $2 < \left| t_{a^*} \right| \leq 3$ ($2 < \left| t_{b^*} \right| \leq 3$), то это свидетельствует о значимости параметра регрессии: надежность в этом случае колеблется от 0,95 до 0,99;

• наконец, если $\left| t_{a^*} \right| > 3$ ($\left| t_{b^*} \right| > 3$), то это почти полная гарантия значимости параметра регрессии. ◂

2.5.2. ОЦЕНКА ЗНАЧИМОСТИ УРАВНЕНИЯ РЕГРЕССИИ

После проверки значимости каждого параметра модели обычно проверяется общее качество уравнения регрессии, которое оценивается по тому, насколько хорошо согласуется уравнение регрессии со статистическими данными. Проверить *значимость уравнения регрессии* – это значит установить, соответствует ли эконометрическая модель, выражающая зависимость между переменными, экспериментальным данным.

Мерой соответствия уравнения регрессии статистическим данным является *коэффициент детерминации* R^2, определяемый по формуле

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (y_i - y_i^*)^2}{\sum (y_i - \bar{y})^2}. \tag{2.27}$$

Величина R^2 показывает, какая часть (доля) вариации (изменения относительно среднего значения) зависимой переменной объясняется уравнением регрессии, а, следовательно, и вариацией объясняющей переменной.

Исходя из определения (2.27), можно показать, что $0 \leq R^2 \leq 1$.

Чем ближе R^2 к единице, тем теснее точки корреляционного поля примыкают к линии регрессии, тем лучше регрессия *аппроксимирует* эмпирические данные. Если $R^2 = 1$, то эмпирические точки (x_i, y_i), $i = 1,2,\ldots, n$, лежат на линии регрессии и между переменными X и Y существует линейная функциональная зависимость: $Y = aX + b$, $a \neq 0$. При ослаблении зависимости между переменными величина R^2 приближается к нулю. Если $R^2 = 0$, то вариация зависимой переменной полностью обусловлена воздействием случайных факторов, и линия регрессии параллельна оси абсцисс: $y^* = \bar{y}$.

Однако не следует абсолютизировать высокое значение R^2. Коэффициент детерминации может быть близким к единице просто в силу того, что обе исследуемые величины X и Y имеют выраженный, например, временной тренд (т.е. зависимость от времени), не связанный с их причинно-следственной зависимостью. В экономике обычно такой тренд имеют объемные показатели (ВНП, ВВП, доход, потребление и т.п.). Поэтому при оценивании регрессий по временным рядам объемных показателей (например, зависимость потребления от дохода или спроса от цены) величина R^2 может быть весьма близкой к единице. Но это не обязательно свидетельствует о наличии значимой линейной связи между исследуемыми показателями, а может означать лишь то, что поведение зависимой переменной нельзя описать уравнением $y = \bar{y}$.

Проверка значимости (или *адекватности* эмпирическим данным) уравнения регрессии осуществляется с помощью F-статистики

$$F=\frac{R^2(n-2)}{1-R^2}. \qquad (2.28)$$

Если справедлива нулевая гипотеза H_0 об отсутствии линейной зависимости между переменными X и Y, то статистика (2.28) имеет F-распределение Фишера с $n_1 = 1$ и $n_2 = n - 2$ степенями свободы. Следовательно, на основании F-критерия Фишера с надежностью $\gamma = 1 - \alpha$ принимается конкурирующая гипотеза H_1 о наличии значимой линейной зависимости между переменными X и Y, если выполняется неравенство

$$F \geq F_{1-\alpha}(n_1, n_2),$$

где $F_{1-\alpha}(n_1, n_2)$ – квантиль порядка $(1 - \alpha)$ F-распределение Фишера с $n_1 = 1$ и $n_2 = n - 2$ степенями свободы.

Пример 2.4. По данным примера 2.2 оценить с надежностью 0,95 значимость параметров и уравнения построенной регрессии.

Решение. По формулам (2.25) и (2.26), используя результаты примеров 2.2 и 2.3, вычислим t-статистики:

$$t_{a^*}=\frac{0{,}9339}{0{,}0387}=24{,}132; \quad t_{b^*}=\frac{3{,}699}{4{,}868}=0{,}76.$$

Уровень значимости $\alpha = 1 - 0{,}95 = 0{,}05$, табличное значение квантиля порядка $1 - \alpha/2 = 0{,}975$ равно $t_{1-\alpha/2}(n-2) = t_{0,975}(10) = 2{,}228$.

Поскольку $\left|t_{a^*}\right| > t_{1-\alpha/2}(n-2)$ (24,132 > 2,228), то нулевая гипотеза H_0: $a = 0$ должна быть отвергнута в пользу альтернативной на уровне значимости 0,05. Это подтверждает статистическую значимость коэффициента регрессии.

Так как $\left|t_{b^*}\right| < t_{1-\alpha/2}(n-2)$ (0,76 < 2,228), то гипотеза о статистической незначимости коэффициента b не отклоняется. Это означает, что в данном случае свободным членом уравнения регрессии можно пренебречь, рассматривая регрессию как $y = ax$.

Рассчитаем коэффициент детерминации по формуле (2.27). Для этого нам понадобится значение $\sum e_i^2 = 35{,}3$ (из примера 2.3). Кроме того, нам потребуется величина $\sum(y_i - \bar{y})^2$, которую вычислим аналогично тому, как нашли $\sum(x_i - \bar{x})^2$ в примере 2.3:

$$\sum_{i=1}^{12}(y_i-\bar{y})^2=\sum_{i=1}^{12}y_i^2-12\cdot(\bar{y})^2=176834-12\cdot\left(\frac{1448}{12}\right)^2=2108{,}667.$$

Здесь мы воспользовались табл. 2.2. Тогда, в силу (2.27), имеем:

$$R^2=1-\frac{35{,}3}{2\,108{,}667}=0{,}983.$$

По формуле (2.28) вычисляем значение F-статистики:

$$F=\frac{0{,}983\cdot(12-2)}{1-(0{,}983)^2}=291{,}596.$$

По таблице 3 Приложения квантилей F-распределения находим

$F_{1-\alpha}(1, n-2) = F_{0,95}(1,10) = 4{,}96$. Так как $F > F_{0,95}(1,10)$ (291,596 > 4,96), то построенное уравнение регрессии значимо.

Отметим, что большое значение коэффициента детерминации (следовательно, и F-статистики) свидетельствует о высоком общем качестве построенной модели зависимости переменных X и Y. ■

Замечание 2.3. Для коэффициента детерминации R^2 можно получить другое полезное представление:

$$R^2=\frac{\sum\left(y_i^*-\overline{y}\right)^2}{\sum\left(y_i-\overline{y}\right)^2}. \qquad (2.29)$$

Формула (2.29) вытекает из (2.27), если воспользоваться важным в *дисперсионном анализе* разложением вариации (суммы квадратов отклонений от средних) зависимой переменной на вариацию, обусловленную линейным воздействием независимой переменной (т.е. уравнением регрессии), и вариацию, обусловленную случайным возмущением:

$$\sum\left(y_i-\overline{y}\right)^2=\sum\left(y_i^*-\overline{y}\right)^2+\sum\left(y_i-y_i^*\right)^2. \blacktriangleleft$$

2.6. ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ РЕГРЕССИОННОЙ МОДЕЛИ

Одна из важнейших задач регрессионного анализа состоит в прогнозировании на основе построенной модели развития изучаемого явления. При этом можно рассматривать *точечный* и *интервальный прогноз* значений зависимой переменной Y, т.е. *определять точечные* и *интервальные оценки* зависимой переменной Y, расположенные вне пределов обследованного диапазона значений независимой переменной X.

Пусть требуется оценить прогнозное среднее значение результирующей переменной Y для заданного значения x_p объясняющей переменной X.

Точечный прогноз рассчитывается по выборочному уравнению регрессии:

$$y_p^* = a^* \cdot x_p + b^*. \quad (2.30)$$

Интервальный прогноз для среднего значения результирующей переменной Y при заданном значении x_p объясняющей переменной X с надежностью γ определяется доверительным интервалом (2.23) при $x = x_p$:

прогнозируемое среднее значение результирующей переменной Y при $x = x_p$ с надежностью $\gamma = 1 - \alpha$ накрывается интервалом

$$\left(y_p^*-t_{1-\alpha/2}(n-2)\cdot S_{y_p^*};\ y_p^*+t_{1-\alpha/2}(n-2)\cdot S_{y_p^*}\right), \qquad (2.31)$$

где $y_p^* = a^* \cdot x_p + b^*$ – *точечный прогноз среднего значения зависимой переменной Y*,

$S_{y_p^*}=S\sqrt{\dfrac{1}{n}+\dfrac{\left(x_p-\overline{x}\right)^2}{\sum\left(x_i-\overline{x}\right)^2}}$ – значение выборочного среднеквадратического отклонение (стандартной ошибки) уравнения регрессии в прогнозируемой точке x_p.

Интервальный прогноз для индивидуального значения результирующей переменной Y при заданном значении x_p объясняющей переменной X с надежностью γ определяется доверительным интервалом (2.24) при $x = x_p$:

прогнозируемое значение результирующей переменной Y при $x = x_p$ с надежностью $\gamma = 1 - \alpha$ принадлежит интервалу

$$\left(y_p^*-t_{1-\alpha/2}(n-2)\cdot S_{Y_{x_p}^*};\ y_p^*+t_{1-\alpha/2}(n-2)\cdot S_{Y_{x_p}^*}\right), \qquad (2.32)$$

где $y_p^* = a^* \cdot x_p + b^*$ – точечный прогноз среднего значения зависимой переменной Y;

$S_{y_p^*}=S\sqrt{1+\dfrac{1}{n}+\dfrac{\left(x_p-\overline{x}\right)^2}{\sum\left(x_i-\overline{x}\right)^2}}$ – значение выборочного среднеквадратического отклонение (стандартной ошибки) индивидуальных значений зависимой переменной Y в прогнозируемой точке x_p.

Пример 2.5. По данным примера 2.2 необходимо:

1) найти с надежностью 0,95 интервальные оценки коэффициентов a, b линейной регрессионной модели и дисперсии возмущения σ^2;

2) оценить среднее потребление одним членом семьи при доходе 160 у.е.;

3) найти 95%-ные доверительный интервал для индивидуального значения потребления на одного члена семьи при доходе 160 у.е.

Решение. 1. При решении примера 2.2 было вычислено, что $\sum_{i=1}^{12} x_i^2 = 190617$, $\bar{x} = 125,25$ и $S = \sqrt{3,53} \approx 1,8788$, а при решении примера 2.3 – $\sum_{i=1}^{12} (x_i - \bar{x})^2$ = 2 366,256.

По таблицам 1 и 2 Приложения находим $t_{1-\alpha/2}(n-2) = t_{0,975}(10) = 2,2281$; $\chi^2_{\alpha/2}(n-2) = \chi^2_{0,025}(10) = 2,247$; $\chi^2_{1-\alpha/2}(n-2) = \chi^2_{0,975}(10) = 20,483$.

Искомые доверительные интервалы получим по формулам (2.20) – (2.22):

$$0,9339 - 2,2281 \cdot \frac{1,8788}{\sqrt{2366,256}} < a < 0,9339 + 2,2281 \cdot \frac{1,8788}{\sqrt{2366,256}};$$

$$3,699 - 2,2281 \cdot 1,8788 \cdot \sqrt{\frac{190617}{12 \cdot 2366,256}} < b < 3,699 + 2,2281 \cdot 1,8788 \cdot \sqrt{\frac{190617}{12 \cdot 2366,256}};$$

$$\frac{12 \cdot 3,53}{20,483} < \sigma^2 < \frac{12 \cdot 3,53}{2,247},$$

или

$$0,8478 < a < 1,0196;$$

$$-7,1473 < b < 14,5453;$$

$$2,0681 < \sigma^2 < 18,8518.$$

2. Уравнение регрессии было получено в примере 2.2: $y^* = 0,9339x + 3,699$. Оценкой среднего потребления $y = M(Y/X = x)$ при $x = 160$ является величина $y^*(160)$, которая находится по уравнению регрессии:

$$y^*(160) = 0,9339 \cdot 160 + 3,699 = 153,123 \text{ (у.е.)}.$$

3. Воспользовавшись формулой (2.32), рассчитываем границы интервала, в котором с надежностью 0,95 будет находиться возможное значение потребления при уровне дохода $x = 160$:

$$153,123 \mp 2,2281 \cdot 1,8788 \cdot \sqrt{1 + \frac{1}{12} + \frac{(160 - 125,25)^2}{2366,256}},$$

или

$$(147,838;\ 158,408).$$

Следовательно, при доходе 160 у.е. возможное потребление с вероятностью 0,95 будет находиться в интервале (147,838 у.е.; 158,408 у.е.). ■

2.7. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ И ПРОВЕРКА ЕГО ЗНАЧИМОСТИ

Одним из важнейших элементов эконометрического анализа является установление наличия и тесноты связи между различными показателями (например, между ценой и спросом, доходом и потреблением, инфляцией и безработицей). Обычно анализ начинают с простейшей – линейной зависимости.

Числовой характеристикой, измеряющей степень тесноты линейной статистической связи между случайными переменными X и Y, является *коэффициент корреляции* между X и Y, который обозначается $\rho = \rho_{X,Y}$ и определяется по формуле

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{M\{(X - MX)(Y - MY)\}}{\sqrt{D(X)D(Y)}}.$$

Приведем основные его свойства.

1. Для любых переменных X и Y абсолютная величина коэффициента корреляции не превосходит единицы: $\left|\rho_{X,Y}\right| \leq 1$, или $-1 \leq \rho_{X,Y} \leq +1$.

2. Абсолютная величина коэффициента корреляции равна единице тогда и только тогда, когда между переменными X и Y существует линейная функциональная зависимость, т. е. $Y = aX + b$, где $a \neq 0$ и b – некоторые постоянные величины. При этом $\rho_{X,Y} = 1$, если $a > 0$, и $\rho_{X,Y} = -1$, если $a < 0$.

3. Если переменные X и Y независимы, то коэффициент корреляции между ними равен нулю. Обратное утверждение (из равенства нулю коэффициента корреляции между X и Y следует их независимость) верно лишь для некоторых частных случаев и неверно в общем случае. В том случае, когда $\rho_{X,Y} = 0$, говорят, что переменные X и Y *некоррелированные*; в противном случае они *коррелированные*.

Из этих свойств вытекает смысл $\rho_{X,Y}$, который состоит в том, что коэффициент корреляции характеризует тесноту линейной статистической связи между переменными X и Y: чем ближе $\left|\rho_{X,Y}\right|$ к единице, тем связь сильнее; чем ближе $\left|\rho_{X,Y}\right|$ к нулю, тем связь слабее. Переменные X и Y называются *положительно коррелированными*, если $\rho_{X,Y} > 0$ и *отрицательно коррелированными*, если $\rho_{X,Y} < 0$.

Определение и свойства теоретического коэффициента корреляции $\rho_{X,Y}$ показывают, что изучение линейной статистической зависимости между переменными X и Y имеет смысл лишь тогда, когда величина $\rho = \rho_{X,Y}$ *значима* (или *существенна*), т. е. не очень близка к нулю. Однако эта величина на практике, как правило, неизвестна и может быть лишь оценена с помощью выборочных данных.

Точечной оценкой теоретического коэффициента корреляции $\rho_{X,Y}$ является *выборочный коэффициент корреляции* $r = r_{xy}$, который находится по формуле

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{S_x \cdot S_y}. \tag{2.33}$$

В формуле (2.33)

$\overline{x}$ и $\overline{y}$ – выборочные средние переменных X и Y соответственно;

$S_x^2 = \frac{1}{n}\sum\left(x_i - \overline{x}\right)^2 = \overline{x^2} - \left(\overline{x}\right)^2$ – выборочная дисперсия переменной X;

$S_y^2 = \frac{1}{n}\sum\left(y_i - \overline{y}\right)^2 = \overline{y^2} - \left(\overline{y}\right)^2$ – выборочная дисперсия переменной Y;

$S_x = \sqrt{S_x^2}$ и $S_y = \sqrt{S_y^2}$ – выборочные среднеквадратические (стандартные) отклонения переменных X и Y соответственно;

$\overline{x \cdot y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$ – выборочное среднее переменной $X \cdot Y$.

Исходя из определения (2.33), можно показать, что выборочный коэффициент корреляции также обладает сформулированными выше свойствами 1–3.

Если выборка имеет достаточно большой объем и хорошо представляет генеральную совокупность (репрезентативна), то заключение о тесноте линейной зависимости между переменными, полученное по выборочным данным, в известной мере может быть распространено и на генеральную совокупность. Для достижения этой цели используется критерий, основанный на распределении Стьюдента.

Пусть основная гипотеза H_0 состоит в том, что корреляция между X и Y не значима, т. е. H_0: $\rho_{X,Y} = 0$. Альтернативная гипотеза $H_1 = \overline{H}_0$ состоит в том, что корреляция между X и Y значима. Если справедлива нулевая гипотеза H_0 и объем выборки n достаточно велик, то статистика

$$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2.34}$$

имеет приближенно распределение Стьюдента с $(n-2)$ степенями свободы.

Для заданного уровня значимости α находим по таблице 1 Приложения

$t_{1-\alpha/2}$ $(n-2)$ – квантиль порядка $(1-\alpha/2)$ распределения Стьюдента с $(n-2)$ степенями свободы. Тогда нулевая гипотеза H_0 принимается при выполнении неравенства

$$|\,t_r\,| < t_{1-\alpha/2}(n-2).$$

В противном случае гипотеза H_0 отклоняется в пользу альтернативной H_1.

Если гипотеза H_0 отклоняется, то это означает, что коэффициент корреляции между ними значим. Следовательно, переменные X и Y коррелированны, и можно изучать линейную зависимость между ними.

Замечание 2.4. Для оценки значимости коэффициента корреляции на начальном этапе исследования можно воспользоваться следующей грубой оценкой:

- $|r_{xy}| \le 0{,}3$ – линейная связь отсутствует;
- $0{,}3 < |r_{xy}| < 0{,}7$ – имеется слабая линейная связь;
- $|r_{xy}| \ge 0{,}7$ – имеется сильная линейная связь. ◂

Пример 2.6. По данным примера 2.2 вычислить коэффициент корреляции между переменными X и Y. Значима ли зависимость между X и Y? Проверить гипотезу на уровне значимости 0,01.

Решение. В примере 2.2 были определены

$$\bar{x}=125{,}25;\ \bar{y}=120{,}667;\ \overline{x\cdot y}=15298{,}083;\ \overline{x^2}=15884{,}75.$$

Вычислим далее среднее

$$\overline{y^2}=\frac{1}{12}\sum_{i=1}^{12} y_i^2=\frac{176834}{12}=14736{,}17;$$

дисперсии

$$S_x^2=\overline{x^2}-(\bar{x})^2=15884{,}75-(125{,}25)^2=197{,}188;$$

$$S_y^2=\overline{y^2}-(\bar{y})^2=14736{,}17-(120{,}667)^2=175{,}646;$$

стандартные отклонения

$$S_x=\sqrt{197{,}188}=14{,}042;\ S_y=\sqrt{175{,}646}=13{,}253.$$

По формуле (2.33) находим

$$r_{xy}=\frac{15298{,}083-125{,}25\cdot 120{,}667}{14{,}042\cdot 13{,}253}=0{,}9915.$$

Данное значение коэффициента корреляции позволяет сделать вывод о сильной (прямой) линейной статистической зависимости между рассматриваемыми переменными X и Y.

Проверим гипотезу H_0 : $\rho_{X,Y}=0$ против альтернативной H_1 : $\rho_{X,Y}\neq 0$ на уровне значимости $\alpha=0{,}01$. По формуле (2.34) вычислим статистику

$$t_r=\frac{0{,}9915\cdot\sqrt{12-2}}{\sqrt{1-(0{,}9915)^2}}=24{,}098.$$

С помощью таблицы квантилей распределения Стьюдента определим

$t_{1-\alpha/2}$ $(n-2)$ = $t_{0,\ 995}$ (10) = 3,1693. Поскольку $|\,t_r\,| > t_{1-\alpha/2}(n-2)$ (24,098 > 3,1693), то коэффициент корреляции $r_{x\,y}$ статистически значим. Следовательно, $\rho_{X,Y}$ существенно отличается от нуля и между переменными X и Y существует сильная линейная статистическая зависимость. ■

Замечание 2.5. Сравнивая формулы (2.7) и (2.33) для коэффициентов регрессии и корреляции соответственно, нетрудно заметить, что в линейной модели между ними существует зависимость:

$$r_{xy}=a^*\cdot\frac{S_x}{S_y}.$$

Так, используя результаты вычислений в примерах 2.2 и 2.6, получаем

$$r_{xy} = 0{,}9339 \cdot \frac{14{,}042}{13{,}253} \approx 0{,}99.$$

Некоторое (незначительное) расхождение с величиной, полученной в примере 2.6, вызвано ошибками округлений. ◀

Замечание 2.6. В случае парной линейной регрессионной модели квадрат коэффициента корреляции между зависимой и независимой переменной равен коэффициенту детерминации:

$$r_{x,y}^2 = R^2.$$

Так для данных примера 2.2 $R^2 = 0{,}983 \approx (0{,}9915)^2 = r_{xy}^2$ (см. примеры 2.5 и 2.6); неточности в данном случае связаны с округлением вычислений. ◀

Контрольные вопросы

1. Что такое функция регрессии?
2. Чем регрессионная модель отличается от функции регрессии?
3. Какая регрессионная модель называется линейной?
4. Какой смысл имеют коэффициенты парного линейного уравнения регрессии?
5. В чем состоит различие между теоретическим и выборочным уравнением регрессии?
6. В чем сущность метода наименьших квадратов (МНК)?
7. Приведите формулы расчета коэффициентов парного линейного уравнения регрессии по МНК.
8. Перечислите предпосылки регрессионного анализа.
9. Сформулируйте основные свойства МНК-оценок.
10. Имеют ли коэффициенты парной линейной регрессии размерность?
11. Как оценивается дисперсия возмущений?
12. Какие факторы влияют на величину стандартных ошибок коэффициентов регрессии?
13. Как строятся интервальные оценки коэффициентов регрессии?
14. Как строятся доверительные полосы для:
    а) уравнения регрессии;
    б) индивидуальных значений результирующей переменной?
15. Какие виды прогнозов Вы знаете?
16. В чем суть предсказания:
    а) среднего значения;
    б) индивидуальных значений результирующей переменной?
17. Объясните суть коэффициента корреляции.
18. Сформулируйте основные свойства коэффициента корреляции.
19. Почему коэффициент корреляции называют мерой линейной зависимости между переменными?
20. В чем суть значимости коэффициента корреляции и как она проверяется?
21. Опишите «грубое» правило анализа статистической значимости коэффициента корреляции.
22. Как связаны коэффициенты регрессии и корреляции в парной
23. регрессионной линейной модели?
24. В чем суть статистической значимости коэффициентов регрессии? Как она проверяется?
25. Опишите «грубое» правило анализа статистической значимости коэффициентов регрессии.
26. Объясните суть коэффициента детерминации.
27. В чем суть статистической значимости уравнения регрессии? Как она проверяется?
28. Как связаны коэффициенты детерминации и корреляции в парной регрессионной линейной модели?

Контрольные задания

1. Какое из следующих утверждений истинно, ложно, не определено? Почему?

а) случайное возмущение ε_i и отклонение e_i совпадают;

б) в регрессионной модели объясняющая переменная является фактором изменения зависимой переменной;

в) коэффициенты теоретического и выборочного уравнений регрессии являются по сути случайными величинами;

г) значения объясняющей переменной парного линейного уравнения регрессии являются случайными;

д) коэффициент регрессии *a* парного линейного уравнения регрессии показывает процентное изменение зависимой переменной *Y* при однопроцентном изменении *X*;

е) коэффициент *a* регрессии *Y* на *X* имеет тот же знак, что и коэффициент корреляции r_{xy}.

2. Суть МНК состоит в:

а) минимизации суммы квадратов коэффициентов регрессии;

б) минимизации суммы квадратов значений зависимой переменной;

в) минимизации суммы квадратов отклонений точек наблюдений от уравнения регрессии;

г) минимизации суммы квадратов отклонений точек теоретического уравнения регрессии от точек выборочного уравнения регрессии.

Выберите правильные ответы.

1. Как Вы считаете, если по одной и той же выборке рассчитаны регрессии *Y* на *X* и *X* на *Y*, то совпадут ли в этом случае линии регрессии?

2. Если переменная *X* принимает значение, равное среднему по выборке $\bar{x}$, то:

а) наблюдаемая величина зависимой переменной *Y* равна среднему значению $\bar{y}$;

б) рассчитанное по выборочному уравнению регрессии значение переменной *Y* в среднем равно $\bar{y}$, но не обязательно равно ему в каждом конкретном случае;

г) рассчитанное по выборочному уравнению регрессии значение переменной *Y* равно среднему значению $\bar{y}$.

Какой из выводов Вам представляется верны и почему?

1. Какое из указанных утверждений истинно, ложно, не определено? Почему?

а) предпосылки регрессионного анализа являются обязательным условием построения линейной регрессионной модели;

б) теоретическим обоснованием МНК является теорема Гаусса-Маркова;

в) оценки коэффициентов регрессии будут иметь нормальное распределение, если случайные возмущения распределены нормально;

г) в любой линейной регрессионной модели, построенной по МНК, справедлива формула $\sum e_i = 0$;

д) построение интервальных оценок для коэффициентов регрессии основано на том, что эти оценки имеют нормальное распределение;

е) для парной линейной регрессии коэффициент корреляции превосходит коэффициент детерминации.

1. По наблюдениям за 50 фирмами в отрасли стремятся построить регрессионную модель $Y = aX + b + \varepsilon$ и оценить коэффициенты *a* и *b* по МНК. Здесь *X* – прибыль фирмы, *Y* – затраты на обновление основного капитала.

а) если прибыль у всех фирм будет одинаковой, возможно ли построение модели?

б) если прибыль фирм не имеет нормального распределения, то использование МНК нецелесообразно? (Да; нет; нет определенного ответа).

в) при оценке коэффициента регрессии *a* по МНК получено завышенное значение a'^*. Какая оценка в этом случае более вероятна для коэффициента *b*: завышенная, заниженная или несмещенная?

2. Стандартная ошибка оценки a^* коэффициента линейной регрессии равна $a^*/2$. Можно ли в этом случае говорить о наличии зависимости *Y* от *X*? Если можно, то оцените степень тесноты этой зависимости.

3. Имеются следующие данные об уровне механизации работ *X* (в %) и производительности труда *Y* (в т/ч) для 14 однотипных предприятий:

x_i	32	30	36	40	41	47	56	54	60	55	61	67	69	76
y_i	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Необходимо:

а) оценить тесноту и направление связи между переменными с помощью коэффициента корреляции;

б) построить корреляционное поле и на основе его визуального анализа сформулировать модель зависимости между X и Y;

в) найти уравнение регрессии Y на X;

г) найти коэффициент детерминации и пояснить его смысл;

д) проверить значимость коэффициентов и уравнения регрессии на 5%-уровне;

е) определить среднюю производительность труда на предприятиях с уровнем механизации работ 60%;

ж) оценить с надежностью 0,95 среднюю производительность труда на предприятиях с уровнем механизации работ 60%;

з) построить аналогичный доверительный интервал для индивидуальных значений производительность труда на тех же предприятиях.

4. При исследовании корреляционной зависимости между ценой на нефть X и индексом нефтяных компаний Y получены следующие данные (в усл. ед.):

$$\bar{x}=45{,}8;\ \bar{y}=4500; S_x^2=4;\ S_y^2=500;\ \overline{x\cdot y}=206135\ .$$

Необходимо:

а) составить уравнение линейной регрессии Y на X;

б) используя уравнение регрессии, найти среднее значение индекса при цене на нефть 50 усл. ед.

5. По данным обследования 30 предприятий получено следующее уравнение регрессии между оценкой Y (млн. усл. ед.) и фактической стоимостью X (млн. усл. ед.) этих предприятий: $y^* = 0{,}875x + 295$. Дать с надежностью 95% прогноз для среднего и индивидуального значений оценки предприятий, фактическая стоимость которых составила 1300 млн. усл. ед., если коэффициент корреляции между X и Y равен 0,76, а среднее квадратическое отклонение переменной X равно 270 млн. усл. ед.

Литература

1. *Доугерти К.* Введение в эконометрику. – М.: ИНФРА-М, 1999. – Гл. 1,2.
2. *Воронович Н.В.*, *Русин Г.Л.* Эконометрика: Методические указания по выполнению контрольных работ. – Новосибирск, НГУЭУ, 2005.
3. *Практикум* по эконометрике / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001. – I раздел.
4. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – Гл. 2.

Тема 3. ОБЩАЯ ЛИНЕЙНАЯ ЭКОНОМЕТРИЧЕСКАЯ МОДЕЛЬ

Социально-экономические явления и процессы, как правило, определяются большим числом одновременно и совокупно действующих факторов. Например, спрос на некоторый товар определяется не только ценой данного товара, но и ценами на заменяющие и дополняющие товары, доходом потребителей и многими другими факторами. В связи с этим возникает задача исследования зависимости одной результирующей переменной Y от нескольких объясняющих переменных X_1, X_2, …, X_m . Эта задача решается с помощью *множественного регрессионного анализа*.

В многомерном случае вместо парной регрессии $M(Y/X = x) = f(x)$ рассматривается *множественная регрессия*

$$M(Y/X_1=x_1, X_2=x_2,..., X_m=x_m)=f(x_1,x_2,..., x_m).$$

Задача оценки статистической взаимосвязи переменных Y и X_1, X_2, …, X_m формулируется аналогично случаю парной регрессии. *Уравнение множественной регрессии* (или *модель взаимосвязи переменных*) может быть представлено в виде

$$Y = f(a, X) + \varepsilon, \tag{3.1}$$

где Y – зависимая (объясняемая) переменная;

$X = (X_1, X_2, \ldots, X_m)$ – вектор независимых (объясняющих) переменных;

a – вектор параметров (подлежащих определению);

ε – случайное возмущение (ошибка, отклонение).

3.1. ЛИНЕЙНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Обобщением парной линейной регрессионной модели является линейная регрессионная модель с m объясняющими переменными (*линейная модель множественной регрессии*):

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_m X_m + \varepsilon. \tag{3.2}$$

Здесь $a = (a_0, a_1, \ldots, a_m)$ – вектор неизвестных параметров размерности $(m + 1)$.

Величина a_j ($j = 1, 2, \ldots, m$) называется j-м *теоретическим коэффициентом регрессии. Он характеризует чувствительность величины Y к изменению X_j. Другими словами, коэффициент регрессии a_j отражает влияние на функцию регрессии*

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m \tag{3.3}$$

объясняющей переменной X_j при условии, что все другие объясняющие переменные модели остаются постоянными: он показывает, на сколько единиц изменяется среднее значение результирующей переменной Y при увеличении объясняющей переменной X_j на одну единицу своего измерения. Величина a_0 – свободный член, определяющий значение функции регрессии (3.3), *когда все объясняющие переменные равны нулю* (*если, конечно, это имеет смысл в рамках модели*).

После выбора линейной функции в качестве модели зависимости рассматриваемых переменных необходимо оценить параметры этой модели.

Предположим, что проведено n наблюдений над объясняющими переменными $X_1, X_2, \ldots, X_m$ и зависимой переменной Y . Обозначим i-е значение зависимой переменной y_i, а объясняющих переменных – $x_{1i}, x_{2i}, \ldots, x_{mi}$. Тогда для каждого наблюдения будет иметь место зависимость вида

$$y_i = a_0 + a_1 \cdot x_{1\,i} + a_2 \cdot x_{2\,i} + \ldots + a_m \cdot x_{m\,i} + \varepsilon_i, \tag{3.4}$$

где ε_i – случайное возмущение i -го наблюдения, $i = 1, 2, \ldots, n$.

Представим данные наблюдений и соответствующие коэффициенты в матричной форме:

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}, \ \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{m1} \\ 1 & x_{12} & x_{21} & \ldots & x_{m1} \\ \ldots & \ldots & \ldots & \ldots & \\ 1 & x_{1n} & x_{21} & \ldots & x_{m1} \end{pmatrix}, \ \boldsymbol{a} = \begin{pmatrix} a_0 \\ a_1 \\ \ldots \\ a_m \end{pmatrix}, \ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}.$$

Здесь $\boldsymbol{Y}$ – n-мерный вектор-столбец значений зависимой переменной Y;

$\boldsymbol{X}$ – матрица размерности $n \times (m + 1)$, в которой $(j + 1)$-й столбец ($j = 1, 2, \ldots, m$) представляет результаты наблюдений независимой переменной X_j (единичный столбец соответствует переменной при свободном члене a_0);

$\boldsymbol{a}$ – вектор-столбец размерности $(m + 1)$ параметров модели;

$\boldsymbol{\varepsilon}$ – вектор-столбец размерности n случайных возмущений (отклонений).

Тогда в матричной форме модель (3.4) примет вид:

$$Y = X \cdot a + \varepsilon. \tag{3.5}$$

Как и в случае парной регрессии, истинные значения параметров a_j ($j = 0, 1, 2, \ldots, m$) по выборке получить невозможно, их можно только оценить. Поэтому определяются

коэффициенты a_j^* **(*j* = 0, 1, 2, …, *m*) так называемого** ***выборочного уравнения множественной регрессии*:**

$$y^* = a_0^* + a_1^* x_1 + a_2^* x_2 + ... + a_m^* x_m \,, \qquad (3.6)$$

где a_j^* – оценка неизвестного параметра a_j (j = 0, 1, 2, …, m).

Самым распространенным методом нахождения оценок параметров множественной линейной регрессии является *метод наименьших квадратов* (МНК). Напомним, что его суть состоит в минимизации суммы квадратов отклонений e_i наблюдаемых значений y_i зависимой переменной Y от значений y_i^* (i = 1, 2, …, n), получаемых по уравнению регрессии (3.6). Эти отклонения рассчитываются, очевидно, по формуле

$$e_i = y_i - y_i^* = y_i - a_0^* - a_1^* x_1 - a_2^* x_2 - ... - a_m^* x_m, \; i = 1, 2, ..., n. \qquad (3.7)$$

Согласно МНК, для нахождения оценок a_j^* (j = 0, 1, 2, …, m) минимизируется функция

$$G = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - a_0^* - a_1^* x_1 - a_2^* x_2 - ... - a_m^* x_m \right)^2 .$$

Необходимым условием минимума функции G является равенство нулю всех ее частных производных по a_j^* (j = 0, 1, 2, …, m). Произведя необходимые вычисления и преобразования, получаем *систему нормальных уравнений* из (m + 1) линейных уравнений с (m + 1) неизвестными:

$$\begin{cases} n \cdot a_0^* + \sum_{i=1}^{n} \sum_{j=1}^{m} a_j^* x_{ji} = \sum_{i=1}^{n} y_i, \\ a_0^* \sum_{i=1}^{n} x_{ji} + \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_j^* x_{ji} \right) x_{ji} = \sum_{i=1}^{n} y_i x_{ji}, \; j = 1, 2, ..., m. \end{cases} \qquad (3.8)$$

Запишем систему (3.8) в матричной форме:

$$\left(\boldsymbol{X}^T \boldsymbol{X} \right) \cdot \boldsymbol{a}^* = \boldsymbol{X}^T \boldsymbol{Y} \,. \qquad (3.9)$$

Здесь $\boldsymbol{X}^T$ – матрица, транспонированная к $\boldsymbol{X}$;

$\boldsymbol{a}^*$ – вектор-столбец размерности (m + 1) оценок a_j^* (j = 0, 1, 2, …, m) параметров модели;

$\boldsymbol{Y}$ – вектор-столбец размерности n значений зависимой переменной Y.

Для решения матричного уравнения (3.9) относительно вектора оценок $\boldsymbol{a}^*$ необходимо ввести следующую *предпосылку* 6^0 (см. п. 2.3, тема 2) для множественного регрессионного анализа:

6^0**.** *Векторы-столбцы матрицы* $\boldsymbol{X}$ *являются линейно независимыми*, *т.е. ранг матрицы* $\boldsymbol{X}$ *равен* (m + 1) *– числу неизвестных параметров*: *rang* ($\boldsymbol{X}$) = m + 1.

Кроме того, предполагается, что число имеющихся наблюдений n больше числа неизвестных параметров, т.е. $n > m + 1$, ибо в противном случае в принципе невозможно получение сколь-нибудь надежных статистических выводов.

Из предположения 6^0 следует, что ранг *симметричной* матрицы $\boldsymbol{X}^T\boldsymbol{X}$, который совпадает с рангом матрицы $\boldsymbol{X}$, равен ее порядку, т.е. *rang* ($\boldsymbol{X}^T\boldsymbol{X}$) = m + 1. Из матричной алгебры известно, что в этом случае матрица $\boldsymbol{X}^T\boldsymbol{X}$ является *невырожденной*, т.е. ее определитель не равен нулю. Следовательно, существует матрица $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, обратная к матрице $\boldsymbol{X}^T\boldsymbol{X}$.

Умножая слева уравнение (3.9) на матрицу $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, получаем

$$\boldsymbol{a}^* = \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{Y} \,. \qquad (3.10)$$

Зная вектор $\boldsymbol{a}^*$, выборочное уравнение множественной регрессии можно записать в виде:

$$y^* = \vec{X}^T \cdot \boldsymbol{a}^* \,, \qquad (3.11)$$

где $\vec{X}^T = (1, x_1, ..., x_m)$ – вектор-строка значений объясняющих переменных X_1,…, X_m (первая координата, равная единице, соответствует свободному члену).

Пример 3.1. Имеются следующие данные о сменной добыче угля на одного рабочего Y (т), мощности пласта X_1 (м) и уровне механизации работ X_2 (%), характеризующие процесс добычи угля в 10 шахтах.

Т а б л и ц а 3.1

i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i
1	8	5	5	6	8	8	6
2	11	8	10	7	9	6	6
3	12	8	10	8	9	4	5
4	9	5	7	9	8	5	6
5	6	7	5	10	12	7	8

Предполагая, что между переменными Y, X_1 и X_2 существует линейная корреляционная зависимость, найти ее аналитическое выражение (уравнение регрессии Y на X_1 и X_2).

Решение. Обозначим

$$\boldsymbol{Y}=\begin{pmatrix}5\\10\\\dots\\8\end{pmatrix},\quad \boldsymbol{X}=\begin{pmatrix}1&8&5\\1&11&8\\\dots&\dots&\dots\\1&12&7\end{pmatrix}$$

(напомним, что в матрицу $\boldsymbol{X}$ вводится дополнительный столбец чисел, состоящий из единиц). Тогда

$$\boldsymbol{X}^T\boldsymbol{X}=\begin{pmatrix}1&1&\dots&1\\8&11&\dots&12\\5&8&\dots&7\end{pmatrix}\cdot\begin{pmatrix}1&8&5\\1&11&8\\\dots&\dots&\dots\\1&12&7\end{pmatrix}=\begin{pmatrix}10&94&63\\94&908&603\\63&603&417\end{pmatrix}$$

и

$$\boldsymbol{X}^T\boldsymbol{Y}=\begin{pmatrix}1&1&\dots&1\\8&11&\dots&12\\5&8&\dots&7\end{pmatrix}\cdot\begin{pmatrix}5\\10\\\dots\\8\end{pmatrix}=\begin{pmatrix}68\\664\\445\end{pmatrix}.$$

Матрицу $\boldsymbol{A}^{-1} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}$ определим по формуле $\boldsymbol{A}^{-1}=\frac{1}{|\boldsymbol{A}|}\widetilde{\boldsymbol{A}}^T$, где $|\boldsymbol{A}|$ – определитель матрицы $\boldsymbol{X}^T\boldsymbol{X}$; $\widetilde{\boldsymbol{A}}$ – матрица, присоединенная к матрице $\boldsymbol{X}^T\boldsymbol{X}$.

Найдем определитель $|\boldsymbol{A}|=|\boldsymbol{X}^T\boldsymbol{X}|$, разложив его по элементам первой строки:

$$|\boldsymbol{A}|=10\cdot(-1)^{1+1}\begin{vmatrix}908&603\\603&417\end{vmatrix}+94\cdot(-1)^{1+2}\begin{vmatrix}94&603\\63&417\end{vmatrix}+63\cdot(-1)^{1+3}\begin{vmatrix}94&908\\63&603\end{vmatrix}= 10\cdot15027 - 94\cdot1209 + 63\cdot(-522) = 3\,738.$$

Вычислим алгебраические дополнения A_{ij} каждого элемента a_{ij} матрицы $\boldsymbol{A}$ по формуле:

$$A_{ij}=(-1)^{i+j}\,\Delta_{ij},$$

где Δ_{ij} – определитель матрицы, полученной из матрицы $\boldsymbol{A}$ вычеркиванием i-й строки и j-го столбца.

Например,

$$A_{11}=(-1)^{1+1}\begin{vmatrix}908&603\\603&417\end{vmatrix}=15027;\ A_{12}=(-1)^{1+2}\begin{vmatrix}94&603\\63&417\end{vmatrix}=-1209;\ A_{13}=(-1)^{1+3}\begin{vmatrix}94&908\\63&603\end{vmatrix}=-522.$$

Составляем присоединенную матрицу

$$\widetilde{\boldsymbol{A}}=(A_{ij})=\begin{pmatrix}15027&-1209&-522\\-1209&201&-108\\-522&-108&244\end{pmatrix}$$

и транспонируем ее:

$$\tilde{A}^T = \left(A_{ji}\right) = \begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}.$$

Отметим, что $\tilde{A}^T = \tilde{A}$, так как матрица $\tilde{A}$ симметричная.

Наконец, получаем

$$\boldsymbol{A}^{-1} = (\boldsymbol{X}^T\boldsymbol{X})^{-1} = \frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}.$$

В соответствии с формулой (3.10), умножая эту матрицу на вектор $\boldsymbol{X}^T\boldsymbol{Y}$, находим

$$\boldsymbol{a}^* = \frac{1}{3738}\begin{pmatrix} -13230 \\ 3192 \\ 1372 \end{pmatrix} = \begin{pmatrix} -3{,}5393 \\ 0{,}8539 \\ 0{,}3670 \end{pmatrix}.$$

С учетом (3.6) выборочное уравнение множественной регрессии имеет вид:

$$y^* = -\,3{,}54 + 0{,}854x_1 + 0{,}367x_2\,.$$

Оно показывает, что при увеличении только мощности пласта на 1 м (при неизменном уровне механизации работ) добыча угля на одного рабочего увеличится в среднем на 0,854 т, а при увеличении только уровня механизации работ на 1% (при неизменной мощности пласта) – в среднем на 0,367 т. ■

3.2. ПРЕДПОСЫЛКИ МНОЖЕСТВЕННОГО ЛИНЕЙНОГО РЕГРЕССИОННОГО АНАЛИЗА. СВОЙСТВА МНК-ОЦЕНОК

Приведенные ранее (см. п. 2.3, тема 2) предпосылки 1^0 – 5^0 для парного регрессионного анализа и, кроме того, предпосылка 6^0 о невырожденности матрицы значений объясняющих переменных для множественного регрессионного анализа могут быть записаны следующим образом:

1^0. В модели (3.5) $\boldsymbol{\varepsilon}$ – случайный вектор, а $\boldsymbol{X}$ – неслучайная (детерминированная) матрица.

2^0. $M(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, где $\boldsymbol{0}$ – нулевой вектор размера n.

3^0, 4^0. Ковариационная матрица вектора возмущений $\boldsymbol{\varepsilon}$

$$\boldsymbol{K(\varepsilon) = M(\varepsilon^T\varepsilon)} = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_n) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_n) \\ \dots & \dots & \dots & \dots \\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \dots & M(\varepsilon_n^2) \end{pmatrix}$$

представима в виде

$$K(\varepsilon) = \sigma^2 E,$$

где $\boldsymbol{E}$ – единичная матрица n-го порядка.

5^0. Вектор возмущений $\boldsymbol{\varepsilon}$ – нормально распределенный случайный вектор, т.е. $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma\boldsymbol{E})$.

6^0. $rang\,(\boldsymbol{X}) = m + 1 < n$.

МНК-оценки (3.10) для модели (3.5) множественной линейной регрессии обладают свойствами, аналогичным свойствам МНК-оценок для парной линейной модели. Другими словами, теорема Гаусса – Маркова, рассмотренная выше для парной регрессионной модели (см. п. 2.3, тема 2), оказывается верной и в общем виде для модели (3.5) множественной регрессии.

***Теорема 3.1* (*Гаусса – Маркова*).** *Если выполнены предпосылки 1^0 – 4^0 и 6^0, то вектор МНК-оценок $\boldsymbol{a}^*$ обладает следующими свойствами*:

- *оценки являются несмещенными, т.е.* $M(\boldsymbol{a}^*) = \boldsymbol{a}$;
- $\mathbf{D}(a_j^*) = \sigma^2 \cdot \mathbf{z_j}$, *(3.12)*

где $\mathbf{z_j} = \mathbf{z_{jj}}$ – элемент матрицы $\boldsymbol{Z} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}$, стоящий на пересечении *j*-й строки и j-го столбца (*j = 0, 1, …, m*);

• **оценки состоятельны:** $a^* \to a$ **по вероятности при** $n \to \infty$**;**

• *оценки эффективны, т.е. они имеют наименьшие дисперсии по сравнению с любыми другими несмещенными оценками, линейными относительно величин* y_i, $i = 1, 2, \dots, n$.

Несмещенной оценкой дисперсии «возмущений» (остаточной дисперсии) σ^2 является *выборочная остаточная дисперсия*

$$S^2 = \frac{1}{n-m-1}\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2 = \frac{1}{n-m-1}\sum_{i=1}^{n} e_i^2, \tag{3.13}$$

где m – количество объясняющих переменных модели;

y_i^* – значение зависимой переменной Y, найденное по выборочному уравнению регрессии (3.6);

$e_i = y_i - y_i^*$ – *выборочная оценка возмущения* ε_i, или *остаток регрессии*.

Оценками дисперсий выборочных коэффициентов регрессии, в силу (3.12), являются величины

$$S_{a_j^*}^2 = S^2 \cdot z_j = \frac{\sum e_i^2}{n-m-1} z_j,\; j = 0, 1, \dots, m. \tag{3.14}$$

Как и в случае парной регрессии, величина $S = \sqrt{S^2}$ часто называется *стандартной ошибкой регрессии*, а $S_{a_j^*} = \sqrt{S_{a_j^*}^2}$ – стандартной ошибкой *j*-го коэффициента регрессии.

Пример 3.2. По данным примера 3.1 вычислить выборочную остаточную дисперсию и стандартные ошибки коэффициентов регрессии.

Решение. Для вычисления выборочной остаточной дисперсии S^2 составим следующую таблицу:

Т а б л и ц а 3.2

i	x_{1i}	x_{2i}	y_i	y_i^*	$e_i^2 = (y_i - y_i^*)^2$
1	8	5	5	5,13	0,016
2	11	8	10	8,79	1,464
3	12	8	10	9,64	1,127
4	9	5	7	5,98	1,038
5	8	7	5	5,86	0,741
6	8	8	6	6,23	0,052
7	9	6	6	6,35	0,121
8	9	4	5	5,61	0,377
9	8	5	6	5,13	0,762
10	12	7	8	9,28	1,631
$\sum$	94	63	68	–	6,329

По формуле (3.13) определяем, что

$$S^2 = \frac{6{,}329}{10-2-1} = 0{,}904 \text{ и } S = \sqrt{0{,}904} = 0{,}951 \text{ (т).}$$

Из найденной в примере 3.1 матрицы $\boldsymbol{Z} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}$ видно, что

$$z_0 = \frac{15027}{3738} = 4{,}02;\; z_1 = \frac{201}{3738} = 0{,}054;\; z_2 = \frac{244}{3738} = 0{,}065.$$

Следовательно, в силу (3.14) выборочные дисперсии и стандартные ошибки коэффициентов регрессии равны:

$$S_{a_0^*}^2 = 0{,}904 \cdot 4{,}02 = 3{,}634;\; S_{a_1^*}^2 = 0{,}904 \cdot 0{,}054 = 0{,}049;\; S_{a_2^*}^2 = 0{,}904 \cdot 0{,}065 = 0{,}059;$$

$$S_{a_0^*} = 1{,}906;\; S_{a_1^*} = 0{,}221;\; S_{a_2^*} = 0{,}243. \blacksquare$$

3.3. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ВО МНОЖЕСТВЕННОЙ РЕГРЕССИИ

По аналогии с парной регрессией (тема 2) после определения точечных оценок a_j^* неизвестных коэффициентов a_j (j = 0, 1, 2, …, m) теоретического уравнения регрессии могут быть рассчитаны интервальные оценки указанных коэффициентов. Аналогичным образом мы можем поступить относительно остаточной дисперсии, уравнения регрессии и отдельных значений результирующей переменной.

3.3.1. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ ПАРАМЕТРОВ МОДЕЛИ

Построение доверительных интервалов для теоретических коэффициентов регрессии a_j и дисперсии «возмущений» σ^2 множественной регрессионной модели (3.5), соответствующих заданной надежности (доверительной вероятности) $\gamma = 1 - \alpha$, опирается на следующее утверждение.

Теорема 3.2. *Если выполнены предпосылки 1^0 – 6^0 множественной линейной регрессионной модели, то*:

• *статистика*

$$t_{a_j^*} = \frac{a_j^* - M(a_j^*)}{S_{a_j^*}} = \frac{a_j^* - a_j}{S \cdot z_j},\ j = 0, 1, ..., m,$$

имеет t -распределение Стьюдента с ($n - m - 1$) *степенями свободы*;

• *статистика*

$$\chi^2 = \frac{\sum e_i^2}{\sigma^2} = \frac{(n-m-1)S^2}{\sigma^2}$$

имеет χ^2 – распределение с ($n - m - 1$) *степенями свободы*.

Пусть $t_{1-\alpha/2}(n-m-1)$ – квантиль порядка ($1 - \alpha/2$) распределения Стьюдента с ($n - m - 1$) степенями свободы;

$\chi^2_{\alpha/2}(n-m-1)$ и $\chi^2_{1-\alpha/2}(n-m-1)$ – квантили порядков $\alpha/2$ и ($1 - \alpha/2$) соответственно χ^2-распределения с ($n - m - 1$) степенями свободы.

Значения этих величин при известных n, m и $\alpha = 1 - \gamma$ можно найти в Приложении по таблицам 1 и 2 соответственно.

Используя стандартную процедуру построения доверительных интервалов, когда известны распределения соответствующих статистик, получим следующие *интервальные оценки для параметров a_j* и σ^2 с надежностью $\gamma = 1 - \alpha$:

$$a_j^* - t_{1-\alpha/2}(n-m-1) S_{a_j^*} < a_j < a_j^* + t_{1-\alpha/2}(n-m-1) S_{a_j^*} \tag{3.15}$$

где $S_{a_j^*} = S\sqrt{z_j}$ – стандартная ошибка j-го коэффициента регрессии (j = 0, 1, …, m);

$$\frac{(n-m-1)S^2}{\chi^2_{1-\alpha/2}(n-m-1)} < \sigma^2 < \frac{(n-m-1)S^2}{\chi^2_{\alpha/2}(n-m-1)}. \tag{3.16}$$

3.3.2. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ФУНКЦИИ РЕГРЕССИИ

Построение доверительного интервала для функции регрессии, т.е. для условного математического ожидания $y = M(Y / X_1 = x_1, X_2 = x_2, ..., X_m = x_m)$, основывается на следующей теореме.

Теорема 3.3. *Если выполнены предпосылки 1^0 – 6^0 множественной линейной регрессионной модели, то статистика*

$$t_{y^*} = \frac{y^* - M(y^*)}{S_{y^*}} = \frac{y^* - y}{S\sqrt{\vec{X}^T Z \cdot \vec{X}}},$$

где $y^* = a_0^* + a_1^* x_1 + ... + a_m^* x_m$ – выборочное уравнение множественной регрессии, при любых x_1, …, $x_m \in R$ имеет t-распределение Стьюдента с ($n - m - 1$) *степенями свободы*.

Воспользовавшись снова стандартной процедурой построения доверительного интервала, когда известно распределение соответствующей статистики, получим следующую *интервальную оценку уравнения регрессии*

$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m$ надежности $\gamma = 1 - \alpha$:

$$y^* - t_{1-\alpha/2}(n-m-1)\cdot S_{y^*} < y < y^* + t_{1-\alpha/2}(n-m-1)\cdot S_{y^*}\ , \qquad (3.17)$$

где $y^* = \vec{X}^T \cdot \boldsymbol{a}^* = a_0^* + a_1^* x_1 + \ldots + a_m^* x_m$ – выборочное уравнение регрессии;

$S_{y^*} = S\sqrt{\vec{X}^T Z \cdot \vec{X}}$ – стандартная ошибка уравнения регрессии.

Множество доверительных интервалов, содержащихся между *нижней* и *верхней доверительными границами* неравенства (3.17), называется *доверительным множеством* для теоретического уравнения множественной регрессии.

3.3.3. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ИНДИВИДУАЛЬНЫХ ЗНАЧЕНИЙ ЗАВИСИМОЙ ПЕРЕМЕННОЙ

Построенное доверительное множество для уравнения множественной регрессии $y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m$ определяет местоположение теоретической линии регрессии в m-мерном векторном пространстве R^m, но не индивидуальных значений зависимой переменной $Y_x = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m + \varepsilon$, которые могут отклоняться от линии регрессии. Определение доверительного интервала для этих величин основывается на другой теореме.

Теорема 3.4. *Пусть* $Y_x^* = a_0^* + a_1^* x_1 + \ldots + a_m^* x_m + \varepsilon$. *Если выполнены предпосылки* 1^0 – 6^0 *множественной линейной регрессионной модели, то статистика*

$$t_{Y_x^*} = \frac{Y_x^* - M\left(Y_x^*\right)}{S_{Y_x^*}} = \frac{Y_x - y^*}{S\sqrt{1 + \vec{X}^T Z \cdot \vec{X}}}\ ,$$

где $y^* = a_0^* + a_1^* x_1 + \ldots + a_m^* x_m$ – *выборочное уравнение регрессии, при любых* $x_1, \ldots, x_m \in R$ *имеет t-распределение Стьюдента с* $(n - m - 1)$ *степенями свободы.*

Доверительный интервал для отдельных возможных значений зависимой переменной Y строится аналогично п. 3.3.2 и определяется по формуле:

$$y^* - t_{1-\alpha/2}(n-m-1)\cdot S_{y^*} < Y < y^* + t_{1-\alpha/2}(n-m-1)\cdot S_{y^*}\ , \qquad (3.18)$$

где $y^* = \vec{X}^T \cdot \boldsymbol{a}^* = a_0^* + a_1^* x_1 + \ldots + a_1^* x_m$ – выборочное уравнение регрессии;

$S_{Y^*} = S\sqrt{1 + \vec{X}^T Z \cdot \vec{X}}$ – выборочное среднее квадратическое отклонение (стандартная ошибка) индивидуальных значений зависимой переменной;

S – выборочное остаточное среднее квадратическое отклонение.

3.4. АНАЛИЗ КАЧЕСТВА МОДЕЛИ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

Построение выборочного уравнения регрессии является начальным этапом эконометрического анализа. Однако построенное по выборке уравнение регрессии часто оказывается неудовлетворительным по тем или иным характеристикам. Поэтому следующей важнейшей задачей эконометрического анализа является проверка качества уравнения регрессии, которая проводится по следующим направлениям:

- проверка статистической значимости коэффициентов уравнения регрессии;
- проверка общего качества уравнения регрессии.

3.4.1. ПРОВЕРКА СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Статистическая значимость коэффициентов a_j^* (j = 0, 1, …, m) множественной линейной регрессии проверяется с помощью статистики $t_{a_j^*} = \frac{a_j^*}{S_{a_j^*}}$, имеющей t-распределение Стьюдента с (n – m – 1) степенями свободы (теорема 3.2). Поэтому a_j^* значимо отличается от нуля (иначе – гипотеза H_0 о равенстве параметра a_j нулю, т.е. H_0: $a_j = 0$, отвергается) на уровне значимости α (или с надежностью $\gamma = 1 - \alpha$), если

$$\left| t_{a_j^*} \right| = \frac{\left| a_j^* \right|}{S_{a_j^*}} \geq t_{1-\alpha/2}(n-m-1), \tag{3.19}$$

где $t_{1-\alpha/2}(n-m-1)$ – квантиль порядка (1 – α/2) распределения Стьюдента с (n – m – 1) степенями свободы.

В противном случае $\left(\left| t_{a_j^*} \right| < t_{1-\alpha/2}(n-m-1) \right)$ коэффициент a_j^* считается статистически незначимым (статистически близким к нулю). Это означает, что переменную X_j можно считать линейно не связанной с зависимой переменной Y, и ее наличие среди объясняющих переменных не оправдано со статистической точки зрения. Не оказывая сколь-нибудь серьезного влияния на зависимую переменную, она лишь искажает реальную картину взаимосвязи. Поэтому после установления того факта, что коэффициент a_j^* статистически незначим, рекомендуется исключить из уравнения переменную X_j. Это не приведет к существенной потере качества модели, но сделает ее более простой.

Отметим, что зачастую строгую проверку значимости коэффициентов регрессии можно заменить, как и в парной регрессии, простым сравнительным анализом (см. замечание 2.2).

3.4.2. ПРОВЕРКА ОБЩЕГО КАЧЕСТВА УРАВНЕНИЯ РЕГРЕССИИ

После проверки значимости каждого коэффициента регрессии обычно проверяется общее качество уравнения регрессии. Для этой цели используется *множественный коэффициент детерминации* (или просто *коэффициент детерминации*) R^2, который вычисляется по формуле

$$R^2 = 1 - \frac{\sum e_i^2}{\sum \left(y_i - \bar{y} \right)^2}. \tag{3.20}$$

Как и в случае парной регрессии, коэффициент детерминации R^2 характеризует долю вариации зависимой переменной, обусловленной регрессией, или изменчивостью объясняющих переменных: чем ближе R^2 к единице, тем лучше регрессия описывает зависимость между объясняющими и результирующей переменными. Поэтому естественно желание построить регрессию с наибольшим R^2.

Вместе с тем использование только одного коэффициента детерминации R^2 для выбора наилучшего уравнения регрессии может оказаться недостаточным. На практике нередко встречаются ситуации, когда плохая модель регрессии может дать сравнительно высокий коэффициент детерминации R^2.

Недостатком коэффициента детерминации R^2 является то, что он, вообще говоря, увеличивается при добавлении в модель новых объясняющих переменных, хотя это и не обязательно означает улучшение качества модели. В этом смысле предпочтительнее использовать «*исправленный*» *коэффициент детерминации* $\acute{R}^2$, определяемый по формуле

$$\acute{R}^2 = 1 - \frac{n-1}{n-m-1}\left(1 - R^2 \right). \tag{3.21}$$

Из (3.21) следует, что $\acute{R}^2 < R^2$ для $m > 1$ и чем больше число объясняющих переменных m, тем меньше $\acute{R}^2$ по сравнению с R^2. Другими словами, он корректируется в сторону уменьшения с ростом числа объясняющих переменных.

После проверки индивидуальной статистической значимости каждого из коэффициентов регрессии обычно анализируется совокупная значимость всех коэффициентов. Такой анализ осуществляется на основе проверки гипотезы о *значимости уравнения множественной регрессии* – гипотезы об одновременном равенстве нулю всех коэффициентов регрессии при объясняющих переменных:

$$H_0 : a_1 = a_2 = \ldots = a_m = 0.$$

Если данная гипотеза не отклоняется, то делается вывод о том, что совокупное влияние всех m объясняющих переменных $X_1, X_2, \ldots, X_m$ модели на результирующую переменную Y можно считать статистически незначимым, а общее качество уравнения регрессии невысоким.

Для проверки гипотезы H_0 используется следующая F-статистика:

$$F = \frac{R^2 \cdot (n-m-1)}{\left(1-R^2\right)\cdot m}. \tag{3.22}$$

При выполнении предпосылок 1^0 – 6^0 множественного регрессионного анализа (см. п. 3.2) и при справедливости нулевой гипотезы H_0 статистика F имеет распределение Фишера с $n_1 = m$ и $n_2 = n - m - 1$ степенями свободы. Следовательно, *критерий значимости уравнения множественной регрессии* на уровне α (с надежностью $\gamma = 1 - \alpha$) может быть записан в виде:

$$F \geq F_{1-\alpha}(n_1, n_2),$$

где $F_{1-\alpha}(n_1, n_2)$ – квантиль порядка $(1 - \alpha)$ F-распределения Фишера с $n_1 = m$ и $n_2 = n - m - 1$ степенями свободы.

Пример 3.3. По данным примера 3.1 определить множественный коэффициент детерминации и проверить значимость полученного уравнения регрессии Y на X_1 и X_2 с надежностью 0,95.

Решение. Из итоговой строки табл. 3.2 находим $\sum_{i=1}^{10} e_i^2 = 6{,}329, \sum_{i=1}^{10} y_i = 68$, откуда $\bar{y} = \sum_{i=1}^{n} y_i / n = 68/10 = 6{,}8$ (т). Также с помощью табл. 3.2 определим $\sum_{i=1}^{10}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{10} y_i^2 - 10\cdot \bar{y}^2 = 496 - 10\cdot 6{,}8^2 = 33{,}6$. Теперь по (3.20) множественный коэффициент детерминации

$$R^2 = 1 - \frac{6{,}329}{33{,}6} = 0{,}812.$$

Коэффициент детерминации $R^2 = 0{,}812$ свидетельствует о том, что вариация исследуемой зависимой переменной Y – сменной добычи угля на одного рабочего – на 81,2% объясняется изменчивостью включенных в модель объясняющих переменных – мощности пласта X_1 и уровня механизации работ X_2.

По формуле (3.21) вычислим также «исправленный» коэффициент детерминации:

$$\hat{R}^2 = 1 - \frac{9}{7}(1 - 0{,}812) = 0{,}758.$$

Зная $R^2 = 0{,}812$, проверим значимость уравнения регрессии на уровне $\alpha = 1 - \gamma = 0{,}05$. Значение F-статистики критерия по (3.22) равно

$$F = \frac{0{,}812\cdot(10-2-1)}{(1-0{,}812)\cdot 2} = 15{,}12,$$

что больше табличного значения квантиля $F_{1-\alpha}(m, n - m - 1) = F_{0,95}(2,7) = 4{,}74$ (см. табл. 3 Приложения). Следовательно, построенное уравнение регрессии *значимо*, т.е. исследуемая зависимая переменная Y достаточно хорошо описывается включенными в регрессионную модель переменными X_1 и X_2. ■

3.5. ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ МНОЖЕСТВЕННОЙ РЕГРЕССИОННОЙ МОДЕЛИ

Как и в парной регрессии, одной из важнейших задач множественного регрессионного анализа является прогнозирование развития изучаемого явления на основе построенной модели. При этом можно рассматривать как *точечный*, так и *интервальный прогноз* значений результирующей переменной *Y*.

Пусть требуется оценить прогнозное среднее значение результирующей переменной *Y* для заданного вектора $\vec{X}_p^T = (1, x_{1p}, ..., x_{mp})$ значений объясняющих переменных $X_1, X_2, \ldots, X_m$.

Точечный прогноз среднего значения результирующей переменной Y рассчитывается по выборочному уравнению регрессии:

$$y_p^* = \vec{X}_p^T \cdot \boldsymbol{a}^* = a_0^* + a_1^* x_{1p} + ... + a_m^* x_{mp} \,. \tag{3.23}$$

Интервальный прогноз среднего значения результирующей переменной *Y* для заданного вектора $\vec{X}_p^T = (1, x_{1p}, ..., x_{mp})$ значений объясняющих переменных X_1, X_2, …, X_m определяется доверительным интервалом (3.17) при $\vec{X}^T = \vec{X}_p^T$:

прогнозируемое среднее значение результирующей переменной Y с надежностью (доверительной вероятностью) $\gamma = 1 - \alpha$ накрывается интервалом с границами

$$y_p^* \mp t_{1-\alpha/2}(n-m-1) \cdot S\sqrt{\vec{X}_p^T Z \cdot \vec{X}_p} \,, \tag{3.24}$$

где y_p^* – точечный прогноз среднего значения Y.

Интервальный прогноз индивидуального значения результирующей переменной *Y* для заданного вектора $\vec{X}_p^T = (1, x_{1p}, ..., x_{mp})$ значений объясняющих переменных X_1, X_2, …, X_m определяется доверительным интервалом (3.18) при $\vec{X}^T = \vec{X}_p^T$:

прогнозируемое значение результирующей переменной Y с надежностью (доверительной вероятностью) $\gamma = 1 - \alpha$ принадлежит интервалу с границами

$$y_p^* \mp t_{1-\alpha/2}(n-m-1) \cdot S\sqrt{1 + \vec{X}_p^T Z \cdot \vec{X}_p} \,, \tag{3.25}$$

где y_p^* – точечный прогноз среднего значения Y.

Пример 3.4. По данным примера 3.1 необходимо:

1) построить 95%-е доверительные интервалы для коэффициентов регрессии и остаточной дисперсии;

2) оценить сменную добычу угля на одного рабочего для шахт с мощностью пласта 8 м и уровнем механизации работ 6%;

3) спрогнозировать с надежностью 0,95 среднее и индивидуальное значения сменной добычи угля на одного рабочего для шахт с мощностью пласта 8 м и уровнем механизации работ 6%.

Решение. 1) С помощью табл. 1 Приложения при числе степеней свободы $n - m - 1 = 10 - 2 - 1 = 7$ и порядке $1 - \alpha/2 = 1 - 0{,}05/2 = 0{,}975$ находим $t_{0,\,975}(7) = 2{,}3646$. По формуле (3.15), учитывая результаты примеров 3.1 и 3.2, определяем 95%-е доверительные интервалы для теоретических коэффициентов множественной регрессии:

$$-3{,}5393 - 2{,}3646 \cdot 1{,}906 < a_0 < -3{,}5393 + 2{,}3646 \cdot 1{,}906;$$

$$0{,}8539 - 2{,}3646 \cdot 0{,}221 < a_1 < 0{,}8539 + 2{,}3646 \cdot 0{,}221;$$

$$0{,}3670 - 2{,}3646 \cdot 0{,}243 < a_2 < 0{,}3670 + 2{,}3646 \cdot 0{,}243 \,,$$

или

$$8{,}0462 < a_0 < 0{,}9676;$$

$$0{,}3313 < a_1 < 1{,}3765;$$

$$0{,}2105 < a_2 < 0{,}9416.$$

По табл. 2 Приложения находим $\chi^2_{\alpha/2}(n-m-1)=\chi^2_{0,025}(7)=1{,}69$ и $\chi^2_{1-\alpha/2}(n-m-1)=\chi^2_{0,975}(7)=16{,}013$. Учитывая результат примера 3.2, по формуле (3.16) получаем:

$$\frac{(10-2-1)\cdot 0{,}904}{16{,}013} < \sigma^2 < \frac{(10-2-1)\cdot 0{,}904}{1{,}69},$$

или

$$0{,}3952 < \sigma^2 < 3{,}7444 \text{ и } 0{,}6286 < \sigma < 1{,}9350.$$

2) По условию надо оценить среднее значение зависимой переменной Y при $\vec{X}_p^T=(1,8,6)$. В силу (3.23) и результатов примера 3.1 находим:

$$y_p^* = -\,3{,}5391 + 0{,}8539\cdot 8 + 0{,}367\cdot 6 = 5{,}494 \text{ (т)}.$$

3) Для построения доверительных интервалов для среднего и индивидуального значений зависимой переменной Y найдем сначала величину

$$\vec{X}_p^T Z\cdot\vec{X}_p=(1;8;6)\frac{1}{3738}\begin{pmatrix}15027 & -1209 & -522\\ -1209 & 201 & -108\\ -522 & -108 & 244\end{pmatrix}\cdot\begin{pmatrix}1\\8\\6\end{pmatrix}=\frac{1}{3738}(2223;-249;78)\cdot\begin{pmatrix}1\\8\\6\end{pmatrix}=\frac{699}{3738}=0{,}187.$$

Теперь $S_{y_p^*}=0{,}951\sqrt{0{,}187}=0{,}411$ (т) и $S_{Y_p^*}=0{,}951\sqrt{1+0{,}187}=1{,}036$ (т).

Поскольку y_p^* = 5,494 и $t_{1-\alpha/2}(n-m-1)=t_{0,975}(7)=2{,}3646$, по формулам (3.24) и (3.25) найдем:

$$5{,}494 - 2{,}3646\cdot 0{,}411 < y_p < 5{,}494 - 2{,}3646\cdot 0{,}411;$$

$$5{,}494 - 2{,}3646\cdot 1{,}036 < Y_p < 5{,}494 - 2{,}3646\cdot 1{,}036,$$

или

$$4{,}522 < y_p < 6{,}466 \text{ (т)};$$

$$3{,}044 < Y_p < 7{,}944 \text{ (т)}.$$

Итак, средняя сменная добыча угля на одного рабочего для шахт с мощностью пласта 8 м и уровнем механизации работ 6% с надежностью 0,95 «накрывается» доверительным интервалом (4,522; 6,466) (т), а соответствующее индивидуальное значение находится в пределах от 3,044 до 7,944 (т). ■

3.6. ЧАСТНАЯ КОРРЕЛЯЦИЯ

Ранее, в п. 2.5 (тема 2), для оценки тесноты линейной связи между переменными был введен коэффициент корреляции. Однако коэффициенты парной корреляции характеризуют тесноту связи переменных, не принимая во внимание возможного влияния на них других переменных. Поэтому во множественной регрессии возникает задача определения тесноты связи между двумя переменными в «чистом» виде, т.е. при устранении влияния на них одной или нескольких переменных. Таким показателем является частный коэффициент корреляции.

Обозначим для удобства $X_0 = Y$. *Выборочным частным коэффициентом корреляции* (или просто *частным коэффициентом корреляции*) между переменными X_i и X_j (i, j = 0, 1, …, m) при фиксированных значениях остальных (m – 1) переменных модели называется величина

$$r_{ij\bullet 1,2,\ldots,m}=-\frac{R_{ij}}{\sqrt{R_{ii}R_{jj}}}, \tag{3.26}$$

где R_{ij} – алгебраическое дополнение элемента $r_{ij}=r_{x_i,x_j}$ корреляционной матрицы

$$
\boldsymbol{R}=\begin{pmatrix} 1 & r_{01} & r_{02} & \dots & r_{0m} \\ r_{10} & 1 & r_{12} & \dots & r_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m0} & r_{m1} & r_{m2} & \dots & 1 \end{pmatrix}.
$$

В частности, в случае трех переменных (m = 2) из (3.26) следует, что

$$
r_{ij\bullet k}=\frac{r_{ij}-r_{ik}r_{jk}}{\sqrt{\left(1-r_{ik}^2\right)\left(1-r_{jk}^2\right)}}. \qquad (3.27)
$$

Пример 3.5. Исследуется зависимость между следующими показателями работы предприятия легкой промышленности:

$X_0 \equiv Y$ – среднемесячная характеристика качества ткани (в баллах);

X_1 – среднемесячное количество профилактических наладок автоматической линии;

X_2 – среднемесячное число обрывов нити.

По итогам годовой работы 35 однородных предприятий были подсчитаны выборочные парные коэффициенты корреляции r_{ij} (i, j = 0, 1, 2): r_{01} = 0,105; r_{02} = 0,024; r_{12} = 0,996.

Проверка на статистическую значимость, проведенная в соответствии с рекомендациями п. 2.5, свидетельствует об отсутствии статистически значимой парной корреляционной связи между качеством ткани, с одной стороны, и числом профилактических наладок и обрывов нити – с другой, что не согласуется с профессиональными представлениями технолога.

Однако расчет частных коэффициентов корреляции по формуле (3.27) дает другие величины:

$$
r_{01\bullet 2}=\frac{r_{01}-r_{02}\cdot r_{12}}{\sqrt{\left(1-r_{02}^2\right)\left(1-r_{12}^2\right)}}=\frac{0{,}105-0{,}024\cdot 0{,}996}{\sqrt{\left(1-0{,}024^2\right)\left(1-0{,}996^2\right)}}=0{,}907,
$$

$$
r_{02\bullet 1}=\frac{r_{02}-r_{01}\cdot r_{21}}{\sqrt{\left(1-r_{01}^2\right)\left(1-r_{21}^2\right)}}=\frac{0{,}024-0{,}105\cdot 0{,}996}{\sqrt{\left(1-0{,}105^2\right)\left(1-0{,}996^2\right)}}=-0{,}906\,.
$$

Полученные значения «чистых» коэффициентов корреляции уже вполне соответствуют нашим представлениям о естественном характере связей между изучаемыми показателями. ■

Контрольные вопросы

1. Как определяется модель множественной линейной регрессии?
2. Что характеризуют коэффициенты регрессии?
3. В чем суть МНК для построения множественного линейного уравнения регрессии?
4. Опишите алгоритм оценки коэффициентов множественной линейной регрессии по МНК в матричной форме.
5. Перечислите предпосылки множественного регрессионного анализа.
6. Сформулируйте теорему Гаусса-Маркова для множественной регрессии.
7. Что представляет собой и как определяется выборочная остаточная дисперсия?
8. Приведите формулы расчета оценок дисперсий и стандартных ошибок коэффициентов регрессии.
9. Как строятся интервальные оценки коэффициентов регрессии и в чем их суть?
10. Как строятся интервальные оценки для уравнения регрессии и индивидуальных значений зависимой переменной?
11. В чем суть прогнозирования: а) среднего значения; б) индивидуального значения зависимой переменной?
12. Как определяется статистическая значимость коэффициентов регрессии?
13. Как определяется множественный коэффициент детерминации и в чем его суть?
14. Чем «исправленный» коэффициент детерминации отличается от обычного?
15. Как осуществляется анализ статистической значимости уравнения регрессии?
16. В чем заключается отличие частного коэффициента корреляции от парного?
17. Можно ли вычислить частные коэффициенты корреляции, зная парные? Если да, то приведите соответствующие формулы.

Контрольные задания

1. Определите (аргументируя свой ответ), истинны, ложны или являются неопределенными следующие утверждения:

а) МНК является наилучшим методом оценки коэффициентов множественной регрессии;

б) выполнение предпосылок множественного регрессионного анализа является обязательным условием применения МНК;

в) стандартные ошибки коэффициентов регрессии во многом определяются количеством наблюдений;

г) если для уравнения регрессии все t-статистики являются большими, то уравнение регрессии качественное;

д) «исправленный» и обычный коэффициенты детерминации совпадают только в тех случаях, когда $R^2 = 1$ или $R^2 = 0$;

е) если $R^2 = 1$, то статистика $F = 1$; если $R^2 = 0$, то статистика $F = 0$;

ж) если для уравнения регрессии F-статистика большая, то уравнение регрессии качественное;

з) при увеличении количества объясняющих переменных в модели всегда увеличивается: обычный коэффициент детерминации; «исправленный» коэффициент детерминации;

и) коэффициент детерминации является мерой сравнения качества любых регрессионных моделей.

2. Имеются следующие данные о годовых ставках месячных доходов по трем акциям за 6-ти месячный период:

Акция	Доходы по месяцам, %					
A	5,4	5,3	4,9	4,9	5,4	6,0
B	6,3	6,2	6,1	5,8	5,7	5,7
C	9,2	9,2	9,1	9,0	8,7	8,6

Есть основания предполагать, что доходы Y по акции C линейно зависят от доходов X_1 и X_2 по акциям A и B . Необходимо:

а) составить уравнение регрессии Y на X_1 и X_2;

б) дать экономическую интерпретацию коэффициентов полученного уравнения регрессии;

в) проверить значимость коэффициентов регрессии с надежностью 0,95;

г) найти коэффициент детерминации и пояснить его смысл;

д) проверить значимость полученного уравнения регрессии на уровне 5%;

е) оценить с надежностью 0,95 доход по акции C, если доходы по акциям A и B составили соответственно 5,5% и 6%.

3. Для объяснения изменения ВНП за 10 лет строится регрессионная модель с объясняющими переменными – потреблением X_1 и инвестициями X_2. Получены следующие статистические данные (в млр. усл. ед.):

ВНП – Y	14	16	18	20	23	23,5	25	26,5	28,5	30,5
Потребление – X_1	8	9,5	11	12	13	14	15	16,5	17	18
Инвестиции – X_2	1,65	1,8	2,0	2,1	2,2	2,4	2,65	2,85	3,2	3,55

Необходимо:

а) составить уравнение регрессии Y на X_1 и X_2;

б) дать экономическую интерпретацию коэффициентов полученного уравнения регрессии;

в) проверить значимость коэффициентов регрессии с надежностью 0,99;

г) найти коэффициент детерминации и пояснить его смысл;

д) проверить значимость полученного уравнения регрессии на уровне 1%;

е) дать прогнозное значение среднего уровня ВНП через три года, если ожидается, что в это время уровень потребления составит 22 млр. усл. ед., а инвестиций – 3,8 млр. усл. ед.

4. По 20 наблюдениям над переменными Y, X_1 и X_2 получены следующие результаты:

$$\sum x_{1i}=4{,}88;\ \sum x_{1i}^2=2{,}52;\ \sum x_{2i}=26{,}7;\ \sum x_{2i}^2=75{,}15;\ \sum y_i=44{,}7;$$
$$\sum x_{1i}x_{2i}=13{,}75;\sum x_{1i}y_i=22{,}1;\sum x_{2i}y_i=125{,}75;\sum y_i^2=210{,}4;\sum e_i^2=0{,}015.$$

Необходимо:

а) составить уравнение множественной линейной регрессии Y на X_1 и X_2;

б) проверить значимость коэффициентов регрессии с надежностью 0,95;

в) найти коэффициент детерминации и «исправленный» коэффициент детерминации; сравнить их значения;

г) оценить качество построенной модели регрессии на уровне 1%;

д) сделать выводы по модели.

Литература

1. *Доугерти К.* Введение в эконометрику. – М.: ИНФРА-М, 1999. – Гл. 5.
2. *Воронович Н.В.*, *Русин Г.Л.* Эконометрика: Методические указания по выполнению контрольных работ. – Новосибирск, НГУЭУ, 2005.
3. *Практикум* по эконометрике /Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001. – II раздел.
4. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – Гл. 3.

Тема 4. НЕЛИНЕЙНЫЕ ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ

Во многих случаях моделирование экономических зависимостей линейными уравнениями дает вполне удовлетворительный результат и может использоваться для анализа и прогнозирования. Однако в силу многообразия и сложности социально-экономических явлений и процессов ограничиться рассмотрением лишь линейных регрессионных моделей невозможно. Многие экономические зависимости не являются линейными по своей сути, и поэтому их моделирование линейными уравнениями регрессии, безусловно, приведет к неоправданно большим ошибкам.

Так, например, нелинейными оказываются производственные функции (зависимости между объемом произведенной продукции и основными факторами производства – трудом, капиталом и т.п.), функции спроса (зависимость между спросом на товары и их ценой или доходом) и другие.

4.1. НЕЛИНЕЙНЫЕ МОДЕЛИ РЕГРЕССИИ И ИХ ЛИНЕАРИЗАЦИЯ

Различают два класса нелинейных моделей регрессии:

• регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по параметрам;

• регрессии, нелинейные по параметрам.

Примерами нелинейной регрессии по включаемым в нее объясняющим переменным могут служить следующие функции:

• *полиномы разных степеней* – $Y=a+bX+cX^2+\varepsilon$; $Y=a+bX+cX^2+dX^3+\varepsilon$ и т.д.

• *равносторонняя гипербола* – $Y=\dfrac{a}{X}+b+\varepsilon$.

Для оценки параметров нелинейных моделей используются два подхода.

Первый подход основан на *линеаризации* модели и заключается в том, что с помощью подходящих преобразований исходных переменных исследуемую зависимость представляют в виде *линейного* соотношения между *преобразованными* переменными.

Второй подход обычно применяется в случае, когда подобрать соответствующее линеаризирующее преобразование не удается. В этом случае применяются методы *нелинейной оптимизации* на основе исходных переменных.

В рамках первого подхода можно линеаризировать модели как нелинейные по переменным, так и нелинейные по параметрам.

Если модель **нелинейна по переменным**, **но линейна по параметрам**, то введением новых переменных ее можно свести к линейной модели и для оценки параметров новой модели использовать обычный метод наименьших квадратов.

Среди подобных моделей следует назвать хорошо известную в эконометрике *обратную* (или *гиперболическую*) регрессионную модель

$$Y = a / X + b + \varepsilon \, . \tag{4.1}$$

Модель (4.1) может быть использована в микроэкономике для описания связи удельных расходов сырья, материалов, топлива с объемом выпускаемой продукции, времени обращения товара с величиной товарооборота и т.п. Модели вида (4.1) находят применение и в макроэкономике. Классическим ее примером является *модель Филлипса*, характеризующая соотношение между нормой безработицы X и процентом прироста безработицы Y.

Введение в модели (4.1) новой переменной $\tilde{X} = 1/X$, сводит ее к линейной:

$$Y = a\tilde{X} + b + \varepsilon \, . \tag{4.2}$$

Параметры модели (4.2) оцениваются обычным методом наименьших квадратов по формулам (2.7) и (2.8) (тема 2), где данные x_i заменяются на $\tilde{x}_i = 1/x_i$, i = 1, 2, …, n.

Другим примером нелинейных зависимостей может служить взаимосвязь доли расходов на товары длительного пользования и общих сумм расходов или доходов. Математическое описание подобного рода взаимосвязей получило название *моделей Энгеля*. На основе исследования семейных расходов Энгель сформулировал следующее утверждение: с ростом дохода доля доходов, расходуемых на продовольствие, уменьшается; соответственно с увеличением дохода доля доходов, расходуемых на непродовольственные товары, будет возрастать. Однако это увеличение не беспредельно, так как сумма долей на все товары не может быть больше единицы.

Оказалось, что гиперболическая модель (4.1) с параметром $a < 0$ хорошо подходит для описания подобной зависимости. Вместе с тем равносторонняя гипербола не является единственно возможной функцией регрессии для модели Энгеля. Для этой цели можно использовать также *полулогарифмическую* модель

$$Y = a \cdot \ln X + b + \varepsilon \, . \tag{4.3}$$

Данная модель линейна по параметрам a и b и нелинейна по независимой переменной X, однако введением новой переменной $\tilde{X} = \ln X$ она приводится к линейной:

$$Y = a\tilde{X} + b + \varepsilon \, . \tag{4.4}$$

Пример 4.1. Имеются данные о среднемесячном доходе семьи X (в тыс. усл. ед.) и расходах на товары длительного пользования Y (в % к доходам):

Т а б л и ц а 4.1

Номер семьи	1	2	3	4	5	6
X	1	2	3	4	5	6
Y	10	13,4	15,4	16,5	18,6,	19,1

Найти уравнение регрессии Y на X:

а) полулогарифмическое;

б) гиперболическое.

Представить их графически.

Решение. а) Вводим новую переменную $\tilde{X} = \ln X$ и вычисляем необходимые средние для расчета функции регрессии:

$$\bar{\tilde{x}} = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i = \frac{1}{6}\sum_{i=1}^{6}\ln x_i = \frac{6{,}57925}{6} = 1{,}0965; \quad \bar{y} = \frac{1}{6}\sum_{i=1}^{6}y_i = \frac{93}{6} = 15{,}5;$$

$$\overline{\tilde{x}^2}=\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2=\frac{1}{6}\sum_{i=1}^{6}(\ln x_i)^2=\frac{9{,}4099}{6}=1{,}568;$$

$$\overline{\tilde{x}\cdot y}=\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\cdot y_i=\frac{1}{6}\sum_{i=1}^{6}\ln x_i\cdot y_i=\frac{113{,}239}{6}=18{,}873\,.$$

Формулы (2.7) и (2.8) (тема 2) дают нам следующие оценки параметров функции регрессии:

$$a^*=\frac{\overline{\tilde{x}\cdot y}-\overline{\tilde{x}}\cdot\overline{y}}{\overline{\tilde{x}^2}-\overline{\tilde{x}}^2}=\frac{18{,}873-1{,}0965\cdot 15{,}5}{1{,}568-1{,}0965^2}=5{,}1335;$$

$$b^*=\overline{y}-a^*\cdot\overline{\tilde{x}}=15{,}5-5{,}1335\cdot 1{,}0965=9{,}871\,.$$

Таким образом, полулогарифмическая функция Энгеля регрессии доли расходов на товары длительного пользования на среднемесячные доходы семьи имеет вид: $y^* = 9{,}871 + 5{,}1335\cdot \ln x$.

График этой функции изображен на рис. 4.1 сплошной линией.

б) Вводим новую переменную $\tilde{X}=1/X$ и вычисляем необходимые средние:

$$\overline{\tilde{x}}=\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i=\frac{1}{6}\sum_{i=1}^{6}(1/x_i)=\frac{2{,}45}{6}=0{,}408;\quad \overline{y}=\frac{1}{6}\sum_{i=1}^{6}y_i=\frac{93}{6}=15{,}5;$$

$$\overline{\tilde{x}^2}=\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2=\frac{1}{6}\sum_{i=1}^{6}(1/x_i)^2=\frac{1{,}4914}{6}=0{,}249;$$

$$\overline{\tilde{x}\cdot y}=\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\cdot y_i=\frac{1}{6}\sum_{i=1}^{6}\left(\frac{y_i}{x_i}\right)=\frac{32{,}862}{6}=5{,}477\,.$$

Применение формул (2.7) и (2.8) (тема 2) дает следующие оценки параметров функции регрессии:

$$a^*=\frac{\overline{\tilde{x}\cdot y}-\overline{\tilde{x}}\cdot\overline{y}}{\overline{\tilde{x}^2}-\overline{\tilde{x}}^2}=\frac{5{,}477-0{,}408\cdot 15{,}5}{0{,}249-0{,}408^2}=-10{,}262;$$

$$b^*=\overline{y}-a^*\cdot\overline{\tilde{x}}=15{,}5-(-10{,}262)\cdot 0{,}408=19{,}687\,.$$

Следовательно, гиперболическая функция Энгеля регрессии доли расходов на товары длительного пользования на среднемесячные доходы семьи имеет вид:

$$y^*=19{,}687-\frac{10{,}262}{x}\,.$$

График данной функции представлен на рис. 4.1 разрывной линией.

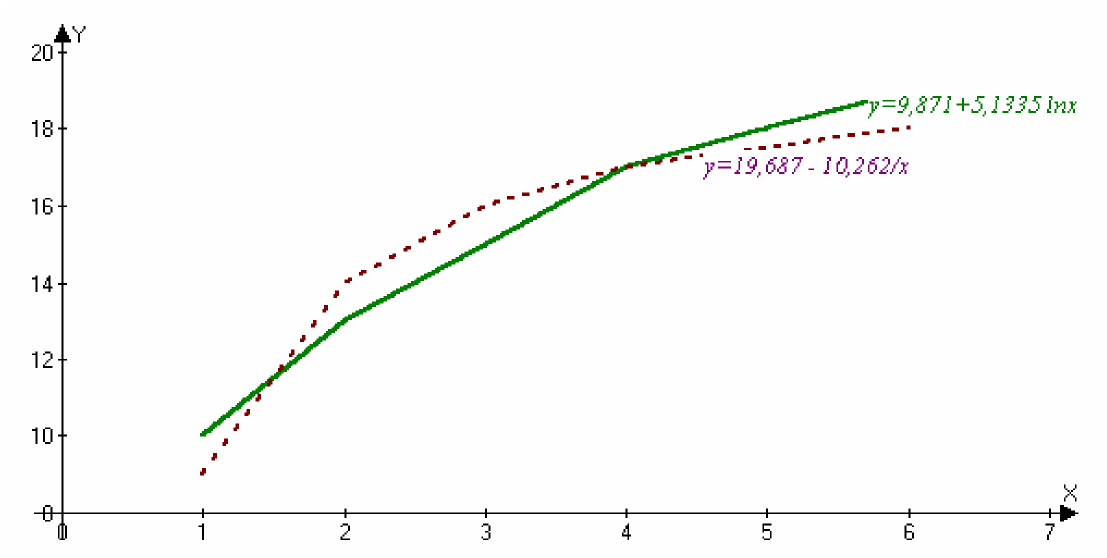


Рис. 4.1. Нелинейные функции регрессии ■

Более сложной проблемой оценивания являются модели **нелинейные по параметрам**, так как непосредственное применение метода наименьших квадратов для оценки параметров таких моделей приводит, как правило, к значительным затруднениям. Примерами подобных нелинейных регрессий могут служить модели, функции регрессий которых есть функции:

- *степенная* $y=a\cdot x^{b}$;
- *показательная* $y=a\cdot b^{x}$;
- *экспоненциальная* $y=e^{a+b\cdot x}$.

В ряде случаев с помощью подходящих преобразований эти модели удается свести к линейным. Так, приведенные выше модели могут быть сведены к линейным логарифмированием обеих частей уравнений.

В качестве примера рассмотрим степенную функцию регрессии

$$y=a\cdot x^{b}, \tag{4.5}$$

которая широко используется в эконометрических исследованиях при изучении зависимости спроса от цены. Логарифмируя обе части данного уравнения, получим

$$\ln y=\ln a+b\cdot\ln x.$$

Путем замены переменных и параметров

$$\tilde{y}=\ln y,\ \tilde{x}=\ln x,\ \tilde{a}=b,\ \tilde{b}=\ln a, \tag{4.6}$$

степенную функцию регрессии можем записать как линейную относительно преобразованных переменных и параметров:

$$\tilde{y}=\tilde{a}\cdot\tilde{x}+\tilde{b}. \tag{4.7}$$

Применение формул (2.7) и (2.8) (тема 2) дает нам следующие оценки параметров функции регрессии (4.7):

$$\tilde{a}^{*}=\frac{\overline{\tilde{x}\tilde{y}}-\overline{\tilde{x}}\cdot\overline{\tilde{y}}}{\overline{\tilde{x}^{2}}-\overline{\tilde{x}}^{2}},\quad \tilde{b}^{*}=\overline{\tilde{y}}-\tilde{a}^{*}\cdot\overline{\tilde{x}},$$

где

$$\overline{\tilde{x}}=\frac{1}{n}\sum_{i=1}^{n}\ln x_i;\ \overline{\tilde{y}}=\frac{1}{n}\sum_{i=1}^{n}\ln y_i;\ \overline{\tilde{x}^{2}}=\frac{1}{n}\sum_{i=1}^{n}(\ln x_i)^2;\ \overline{\tilde{x}\tilde{y}}=\frac{1}{n}\sum_{i=1}^{n}\ln x_i\cdot\ln y_i.$$

Поскольку нам нужны оценки параметров a и b исходной функции регрессии (4.5), то для их нахождения производим преобразования, обратные преобразованиям (4.6):

$$a^{*}=e^{\tilde{b}^{*}},\ b^{*}=\tilde{a}^{*}.$$

Замечание 4.1. Следует отметить существенный недостаток процедуры линеаризации, сводящей нелинейные модели к линейным. Это связано с тем, что МНК-оценки параметров получаются не из условия минимизации суммы квадратов отклонений для исходных переменных, а из условия минимизации суммы квадратов отклонений для преобразованных переменных, что не одно и тоже. В связи с этим необходимо определенное уточнение свойств полученных оценок. ◂

4.2. ПРОИЗВОДСТВЕННАЯ ФУНКЦИЯ КОББА-ДУГЛАСА

Производственная функция – это экономико-математическая модель, характеризующая зависимость объема выпускаемой продукции от факторов производства. При этом модель может быть построена как для отдельной фирмы или отрасли, так и для всей национальной экономики.

Рассмотрим производственную функцию, включающую два фактора производства, – *производственную функцию Кобба-Дугласа*

$$Y=AK^{\alpha}L^{\beta}, \tag{4.8}$$

где Y – объем производства; K – затраты капитала; L – затраты труда; A, α, β – параметры модели, причем $A > 0$, $0 < \alpha < 1$, $0 < \beta < 1$.

Свойства производственной функции Кобба-Дугласа

Эластичность выпуска продукции.

Параметры α и β являются *коэффициентами частной эластичности* объема производства Y по затратам капитала K и труда L соответственно:

$$\frac{\partial Y}{Y}\Bigg/\frac{\partial K}{K}=\alpha,\ \frac{\partial Y}{Y}\Bigg/\frac{\partial L}{L}=\beta.$$

Они показывают, на сколько процентов в среднем изменится Y, если K и L увеличить соответственно на один процент: увеличение затрат капитала на 1% приведет к росту выпуска продукции на α %, а увеличение затрат труда на 1% приведет к росту выпуска на β %.

2. ***Эффект от масштаба производства.***

При росте затрат каждого из факторов K и L в λ раз выпуск продукции возрастет в $\lambda^{\alpha+\beta}$ раз. Это означает следующее:

• если $\alpha + \beta > 1$, то функция Кобба-Дугласа имеет *возрастающую* отдачу от масштаба производства;

• если $\alpha + \beta < 1$, то функция Кобба-Дугласа имеет *убывающую* отдачу от масштаба производства;

• если $\alpha + \beta = 1$, то функция Кобба-Дугласа имеет *постоянную* отдачу от масштаба производства.

3. ***Прогнозируемые доли производственных факторов.***

В рыночной экономике оценки α и β интерпретируются как прогнозируемые доли дохода, полученные соответственно за счет капитала и труда.

Учитывая влияние случайных возмущений, присущих каждому экономическому явлению, функцию Кобба-Дугласа (4.8) можно представить в виде

$$Y = AK^{\alpha}L^{\beta}\varepsilon\ . \qquad (4.9)$$

Полученную степенную (мультипликативную) модель легко свести к линейной путем логарифмирования обеих частей уравнения (4.9):

$$\ln Y = \ln A + \alpha\ln K + \beta\ln L + \ln\varepsilon\ . \qquad (4.10)$$

Введя новые переменные и параметры

$$\tilde{Y}=\ln Y,\ \tilde{A}=\ln A,\ \tilde{K}=\ln K,\ \tilde{L}=\ln L,\ \tilde{\varepsilon}=\ln\varepsilon\ ,$$

получим модель множественной линейной (относительно преобразованных переменных и параметров) регрессии

$$\tilde{Y}=\tilde{A}+\alpha\tilde{K}+\beta\tilde{L}+\tilde{\varepsilon}\ .$$

Предположим, что в модели (4.9) $\alpha + \beta = 1$. Это условие означает, что при расширении масштаба производства – увеличении затрат капитала K и труда L в некоторое число раз – объем производства возрастет в то же число раз. Тогда функцию Кобба-Дугласа можно представить в виде

$$Y = AK^{\alpha}L^{1-\alpha}\varepsilon\ ,$$

или

$$\frac{Y}{L}=A\left(\frac{K}{L}\right)^{\alpha}\varepsilon\ . \qquad (4.11)$$

Таким образом, получаем зависимость производительности труда (Y/L) от капиталовооруженности (K/L). Для оценки параметров модели (4.11) путем логарифмирования приводим ее к модели парной линейной регрессии:

$$\ln(Y/L)=\ln A+\alpha\ln(K/L)+\ln\varepsilon\ . \qquad (4.12)$$

Функция Кобба-Дугласа с учетом технического прогресса имеет вид:

$$Y = AK^{\alpha}L^{\beta}e^{\theta t}\varepsilon\ , \qquad (4.13)$$

где t – время; параметр θ – темп прироста объема производства благодаря техническому прогрессу. Модель (4.13) приводится к линейному виду аналогично модели (4.9).

Пример 4.2. По данным 50 предприятий легкой промышленности получена оценка производственной функции Кобба-Дугласа в следующем виде:

$$\ln\left(y^*/L\right)=0{,}43+0{,}25\cdot\ln(K/L),$$

или

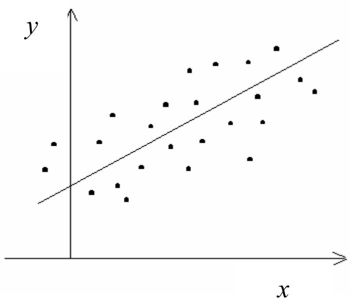
$$\frac{y^*}{L}=1{,}537\left(\frac{K}{L}\right)^{0,25}.$$

Коэффициент регрессии, равный 0,25, который одновременно является коэффициентом эластичности, показывает, что при изменении капиталовооруженности труда на 1% производительность труда на предприятиях отрасли увеличивается в среднем на 0,25%. ■

4.3. ВЫБОР ФОРМЫ МОДЕЛИ

Многообразие и сложность экономических явлений и процессов предопределяет многообразие моделей, используемых для эконометрического анализа. С другой стороны, это существенно усложняет процесс нахождения максимально адекватной модели. Сразу же возникает вопрос: использовать линейную зависимость или нелинейную, и если последнюю, то какого вида?

В случае парной регрессии выбор модели обычно осуществляется по виду расположения наблюдаемых точек на корреляционном поле. Например, экспериментальные данные, изображенные на рис.4.2, явно наводят на мысль о линейной зависимости (модели) вида $Y = aX + b + \varepsilon$. Зависимость, изображенная на рис. 4.3, может быть хорошо представлена многочленом (моделью) второй степени $Y = aX^2 + bX + c + \varepsilon$.



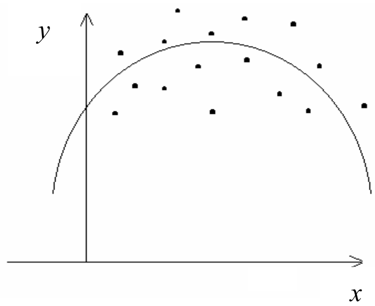


Рис. 4.2. Линейная *Рис. 4.3.* Квадратичная модель модель

Однако нередки ситуации, когда расположение точек приблизительно соответствует нескольким функциям и необходимо выбрать из них наилучшую. Например, нелинейные зависимости могут быть близки к полиномиальной, показательной, степенной, логарифмической функциям. Еще более неоднозначна ситуация для множественной регрессии, где наглядное представление статистических данных невозможно. В таких случаях для выбора вида регрессионной зависимости можно использовать коэффициент детерминации и среднюю ошибку аппроксимации.

4.3.1. КОЭФФИЦИЕНТ ДЕТЕРМИНАЦИИ ДЛЯ НЕЛИНЕЙНОЙ РЕГРЕССИИ

Коэффициент детерминации R^2 для нелинейной регрессии определяется так же, как и для линейной регрессии:

$$R^2=1-\frac{\sum\left(y_i-y_i^*\right)^2}{\sum\left(y_i-\bar{y}\right)^2}, \tag{4.14}$$

где y_i – i-е эмпирическое значение объясняющей переменной Y;

y_i^* – i-е значение объясняющей переменной Y, определенное по уравнению регрессии (i = 1, 2, …, n);

$\bar{y}$ – выборочное среднее переменной Y.

Величина данного показателя находится в пределах от 0 до 1: $0 \le R^2 \le 1$. Как и в линейной регрессии, чем ближе коэффициент детерминации к единице, тем теснее связь рассматриваемых переменных, тем более надежно построенное уравнение регрессии.

Однако, как и в линейной регрессии, не следует абсолютизировать высокое значение R^2. Коэффициент детерминации может быть близким к единице просто в силу того, что исследуемые величины имеют выраженный тренд, не связанный с их причинно-следственной зависимостью.

4.3.2. СРЕДНЯЯ ОШИБКА АППРОКСИМАЦИИ

Фактические значения y_i результирующей переменной Y отличаются, как правило, от значений y_i^*, рассчитанных по уравнению регрессии. Чем меньше это отличие, тем ближе расчетные значения подходят к эмпирическим данным, тем лучше качество модели. Величина отклонений фактических и расчетных значений результирующей переменной $(y_i - y_i^*)$ по каждому наблюдению представляет собой *абсолютную*, а $(y_i - y_i^*)/y_i$ – *относительную ошибку аппроксимации*.

Чтобы иметь общее представление о качестве модели, из относительных ошибок по каждому наблюдению определяют *среднюю ошибку аппроксимации*

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(y_i - y_i^*)}{y_i}\right| \cdot 100\% . \qquad (4.15)$$

Если средняя ошибка аппроксимации находится в пределах 5÷7%, то это свидетельствует о хорошем подборе модели к исходным данным.

Пример 4.3. По данным примера 4.1 с помощью коэффициента детерминации и средней ошибки аппроксимации сравнить полулогарифмическую и гиперболическую модели регрессии и выбрать наилучшую.

Решение. Для удобства вычислений составляем вспомогательные таблицы (табл. 4.2 и табл. 4.3).

Т а б л и ц а 4.2

Полулогарифмическая регрессия $y^* = 9{,}871 + 5{,}1335 \cdot \ln x$

i	x_i	y_i	y_i^*	$y_i - y_i^*$	$\lvert(y_i - y_i^*)/y_i\rvert \cdot 100$	$(y_i - y_i^*)^2$	$(y_i - \bar{y})^2$
1	1	10,0	9,871	0,129	1,29	0,017	0,25
1	2	3	4	5	6	7	8
2	2	13,4	13,429	–0,029	0,22	0,001	4,41
3	3	15,4	15,511	–0,111	0,72	0,012	0,01

Окончание табл. 4.2

1	2	3	4	5	6	7	8
4	4	16,5	16,987	–0,487	2,95	0,237	1,00
5	5	18,6	18,133	0,467	2,51	0,218	9,61
6	6	19,1	19,069	0,031	0,16	0,001	12,96
Σ	21	93,0	93,0	0,0	7,85	0,486	28,23

По формулам (4.14) и (4.15) находим соответственно коэффициент детерминации и среднюю ошибку аппроксимации:

$$R^2 = 1 - \frac{0{,}486}{28{,}23} = 0{,}983; \quad \overline{A} = \frac{7{,}85}{6} = 1{,}31\% .$$

Полученные значения коэффициента детерминации и средней ошибки аппроксимации указывают на хорошее качество уравнения полулогарифмической регрессии.

Т а б л и ц а 4.3

Гиперболическая регрессия $y^* = 19{,}687 - \dfrac{10{,}262}{x}$

i	x_i	y_i	y_i^*	$y_i - y_i^*$	$\lvert(y_i - y_i^*)/y_i\rvert \cdot 100$	$(y_i - y_i^*)^2$	$(y_i - \bar{y})^2$
1	1	10,0	9,43	0,57	5,7	0,325	0,25
2	2	13,4	14,56	–1,16	0,87	1,346	4,41
3	3	15,4	16,28	–0,88	5,71	0,774	0,01

4	4	16,5	17,12	–0,62	3,75	0,384	1,00
5	5	18,6	17,63	0,97	5,21	0,941	9,61
6	6	19,1	17,97	1,13	5,92	1,277	12,96
Σ	21	93,0	93,0	0,0	21,95	5,047	28,23

Привлекая снова формулы (4.14) и (4.15), находим соответственно коэффициент детерминации и среднюю ошибку аппроксимации:

$$R^2 = 1 - \frac{5,047}{28,23} = 0,82;\ \ \overline{A} = \frac{21,95}{6} = 3,66\%.$$

Значения коэффициента детерминации и средней ошибки аппроксимации указывают на достаточно хорошее качество уравнения гиперболической регрессии, хотя его качество уступает полулогарифмической регрессии. ■

Контрольные вопросы

1. Какие регрессионные модели называются нелинейными:
   а) по переменным;
   б) по параметрам?
2. Как оцениваются параметры моделей:
   а) нелинейных по переменным, но линейных по параметрам;
   б) нелинейных по параметрам?
3. Сохраняются ли свойства МНК-оценок (теорема Гаусса-Маркова) для оценок параметров линеаризированных моделей?
4. Можно ли использовать коэффициент детерминации для оценки качества нелинейной модели?
5. Будет ли для нелинейной парной модели коэффициент детерминации совпадать с квадратом коэффициента корреляции?
6. В чем смысл средней ошибки аппроксимации и как она определяется?

Контрольные задания

1. Какие из указанных моделей и каким образом могут быть сведены к линейной модели?

а) $Y = a\ln^2 X + b + \varepsilon$;

б) $Y = \dfrac{a}{\sqrt{X}} + b + \varepsilon$;

в) $Y = \dfrac{1}{aX + b} + \varepsilon$;

г) $Y = \dfrac{X}{aX^2 + b} + \varepsilon$.

2. Предложена следующая регрессионная модель:

$Y = a_0 + a_1 X_1 X_3 + a_2 X_2 X_3 + a_1 a_3 X_1 + a_2 a_3 X_2 + \varepsilon.$

а) является ли данная модель линейной относительно переменных либо параметров?

б) можно ли получить оценки параметров модели при имеющемся наборе данных $(x_{1i}, x_{2i}, x_{3i}, y_i),\ i = 1,2,...,n$?

3. Изучается зависимость стоимости Y (руб.) одного экземпляра книги от тиража X (тыс. экз.) по следующим данным, поученным в результате 8 обследований:

i	1	2	3	4	5	6	7	8
x_i	1	2	3	5	10	20	30	50
y_i	9,1	5,3	4,11	2,83	2,11	1,62	1,41	1,3

Необходимо:

а) найти линейную и гиперболическую функцию регрессии Y на X;

б) построить графики функций регрессии и сравнить их визуально;
в) вычислить коэффициенты детерминации для моделей и по ним выбрать лучшую модель;
г) вычислить средние ошибки аппроксимаций для моделей и по ним выбрать лучшую модель;
д) сравнить, совпадают ли выводы, сделанные в п.п. в) и г);
е) если выводы в п.п. в) и г) не совпадают, то выбрать наилучшую модель.

Литература

1. *Доугерти К.* Введение в эконометрику. – М.: ИНФРА-М, 1999. – Гл. 4.
2. *Практикум* по эконометрике /Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001. – II раздел.
3. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – Гл. 2.

....

Тема 5. ПРОВЕРКА ВЫПОЛНЕНИЯ ОСНОВНЫХ ПРЕДПОСЫЛОК РЕГРЕССИОННОГО АНАЛИЗА

При моделировании реальных экономических явлений и процессов мы нередко сталкиваемся с ситуациями, в которых условия классической линейной модели регрессии (парной или множественной) оказываются нарушенными. Приступая к оценке параметров уравнения регрессии, мы предполагали, что реальная взаимосвязь переменных линейна, а отклонения от регрессионной прямой удовлетворяют определенным условиям. Так ли это на самом деле? Если нет, то наш анализ коэффициентов регрессии неточен и оценки этих коэффициентов могут не обладать такими желательными свойствами, как несмещенность, состоятельность и эффективность.

Для получения качественных оценок необходимо следить за выполнением предпосылок регрессионного анализа (условий Гаусса-Маркова), так как при их нарушении МНК может давать оценки с плохими статистическими свойствами. В частности, могут не выполняться предположения о том, что случайные возмущения модели имеют постоянную дисперсию и не коррелированны между собой.

Заметим, что предположение о нулевом среднем возмущений не является ограничением для модели и вводится для упрощения выкладок. Действительно, пусть в линейной эконометрической модели $Y = aX + b + \varepsilon$ случайное возмущение ε таково, что его среднее $M\varepsilon \neq 0$. Тогда эту модель можно переписать следующим образом:

$$Y = aX + (b + M\varepsilon) + (\varepsilon - M\varepsilon) = aX + \tilde{b} + \tilde{\varepsilon} ,$$

где $\tilde{b} = b + M\varepsilon$ – неизвестный параметр модели (как и параметр b);

$\tilde{\varepsilon} = \varepsilon - M(\varepsilon)$ – случайное возмущение модели с $M\tilde{\varepsilon} = M(\varepsilon - M\varepsilon) = M\varepsilon - M\varepsilon = 0$.

Таким образом, получаем снова линейную модель с неизвестными параметрами, но с нулевым средним для случайного возмущения модели.

Серьезной проблемой при построении множественных эконометрических моделей по МНК является мультиколлинеарность – высокая взаимная коррелированность объясняющих переменных. Изучению указанных проблем посвящена данная тема.

5.1. ГЕТЕРОСКЕДАСТИЧНОСТЬ

5.1.1. СУТЬ И ПОСЛЕДСТВИЯ ГЕТЕРОСКЕДАСТИЧНОСТИ

Одной из существенных предпосылок регрессионного анализа является постоянство дисперсий случайных возмущений ε_i для всех наблюдений (*гомоскедастичность*): $D\varepsilon_i = \sigma^2$, $i = 1, 2,\ldots, n$. Однако на практике это условие нередко нарушается, и мы имеем дело с *гетероскедастичностью* модели. Например, если исследуется зависимость расходов на питание в семье от ее общего дохода, то можно ожидать, что разброс данных будет выше для семей с более высоким доходом. Аналогично, если рассматривается зависимость прибыли предприятия от каких-либо факторов, скажем, от размера основных фондов, то естественно ожидать, что для больших предприятий колебание прибыли будет выше, чем для малых.

При гетероскедастичности линейной регрессионной модели (парной или множественной) **последствия** применения МНК будут следующими.

1. Хотя МНК-оценки останутся по-прежнему несмещенными и линейными, они не будут эффективными, т.е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра. Увеличение дисперсии оценок ведет к уменьшению точности оценок.

2. Стандартные ошибки коэффициентов регрессии будут рассчитываться со смещением. Смещенность возникает вследствие того, что выборочная остаточная дисперсия $S^2 = \sum e_i^2/(n-m-1)$ (m – число объясняющих переменных модели), которая используется при вычислении указанных величин (см. формулы (2.18) и (2.19)), не является более несмещенной.

Вследствие вышесказанного все выводы, получаемые на основе соответствующих t- и F-статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получаемые при проверке качества оценок (параметров модели и самой модели в целом), могут быть ошибочными и приводить к неверным заключениям по построенной модели.

5.1.2. ОБНАРУЖЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

В ряде случаев, зная характер данных, появление проблемы гетероскедастичности можно предвидеть (как в примерах п. 5.1.1) и попытаться устранить этот недостаток еще на этапе спецификации модели. Однако чаще всего эту проблему приходится решать после построения уравнения регрессии.

Обнаружение гетероскедастичности в каждом конкретном случае является довольно сложной задачей, так как для знания дисперсий возмущений необходимо знать закон распределения случайной переменной Y, соответствующий каждому выбранному значению x_i переменной X. На практике часто для каждого конкретного значения x_i определяется единственное значение y_i, что не позволяет оценить дисперсию переменной Y.

Заметим также, что не существует какого-либо однозначного метода определения гетероскедастичности. Однако к настоящему времени для такой проверки известно довольно большое число критериев и тестов для них. Рассмотрим наиболее наглядные и простые.

а) графический анализ остатков

Использование графического представления отклонений фактических значений y_i зависимой переменной Y от значений y_i*, полученных в результате расчета по выборочному уравнению регрессии, позволяет определиться с наличием гетероскедастичности. В этом случае по оси абсцисс откладываются значения x_i объясняющей переменной X, а по оси ординат либо *остатки* $e_i = y_i - y_i^*$, либо их квадраты e_i^2, $i = 1, 2, \dots, n$.

Смысл использования графиков остатков состоит в том, что всякое отклонение от сделанных предположений относительно случайных возмущений ε_i отражается на остатках e_i. Примеры таких графиков приведены на рис 5.1.

На рис. 5.1, *а*) все отклонения e_i^2 находятся внутри полуполосы постоянной ширины, параллельной оси абсцисс. Это позволяет говорить, о независимости дисперсий e_i^2 от значений x_i переменной X и их постоянстве, т.е. в этом случае скорее всего выполняются условия гомоскедастичности модели.

На рис. 5.1, *б*) – *д*) наблюдаются систематические изменения между значениями x_i переменной X и квадратами остатков e_i^2, что отражает большую возможность наличия гетероскедастичности модели.

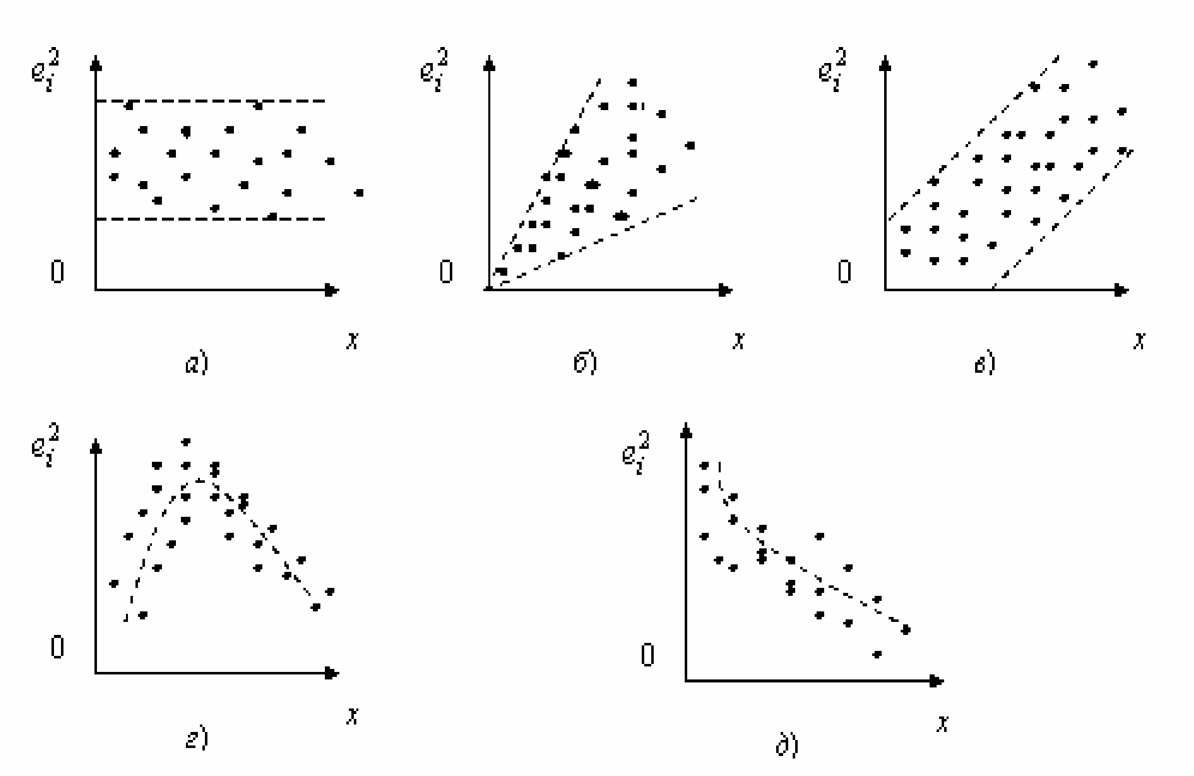


Рис. 5.1. Зависимость остатков от значений объясняющей переменной

Отметим, что графический анализ остатков является удобным инструментом в случае парной регрессии. При множественной регрессии графический анализ возможен для каждой из объясняющих переменных X_j (j = 1, 2, …, m) отдельно. Часто используется также зависимость e_i^2 от значений y_i*, получаемых по выборочному уравнению регрессии. В этом случае вместо значений объясняющих переменных X_j по оси абсцисс откладывают значения y_i*, i = 1, 2, …, n. Поскольку значения y_i* являются линейной комбинацией значений объясняющих переменных X_j (j = 1, 2, …, m), то график, отражающий зависимость e_i^2 от y_i*, будет указывать на наличие гетероскедастичности аналогично ситуациям на рис. 5.1, *б*) – *д*). Такой анализ наиболее целесообразен при большом количестве объясняющих переменных.

б) тест ранговой корреляции Спирмена

Идея данного теста для проверки гетероскедастичности заключается в том, что абсолютные величины остатков e_i являются оценками σ_i , поэтому в случае гетероскедастичности абсолютные величины остатков e_i и значения x_i объясняющей переменной будут коррелированны.

Для использования данного теста все значения x_i и e_i ранжируются (упорядочиваются) по возрастанию:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \text{ и } e_{(1)} \le e_{(2)} \le \ldots \le e_{(n)}.$$

Здесь $x_{(1)} = \min\limits_{i=1,2,\ldots,n} \{x_i\},\ x_{(n)} = \max\limits_{i=1,2,\ldots,n} \{x_i\};\ e_{(1)} = \min\limits_{i=1,2,\ldots,n} \{e_i\},\ e_{(n)} = \max\limits_{i=1,2,\ldots,n} \{e_i\}.$

Затем определяется *коэффициент ранговой корреляции Спирмена*:

$$r_{x,e} = 1 - 6 \cdot \frac{\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{5.1}$$

где $d_i = r_{x_i} - r_{e_i}$ – разность между рангами (номерами в упорядоченной совокупности) величин x_i и e_i , i = 1, 2, … , n;

n – число пар наблюдений.

Если теоретический коэффициент ранговой корреляции Спирмена $\rho_{x,e}$ равен нулю, то статистика

$$t_r = \frac{r_{x,e}\sqrt{n-2}}{\sqrt{1-r_{x,e}^2}} \tag{5.2}$$

при $n \ge 10$ имеет распределение, близкое к t-распределению Стьюдента с $(n - 2)$ степенями свободы. Поэтому при выполнении неравенства

$$|t_r| \ge t_{1-\alpha/2}(n-2),$$

где $t_{1-\alpha/2}(n-2)$ – квантиль порядка $(1 - \alpha/2)$ распределения Стьюдента с $(n - 2)$ степенями свободы необходимо отклонить гипотезу о равенстве нулю коэффициента ранговой корреляции $\rho_{x,e}$, а, следовательно, и об отсутствии гетероскедастичности на уровне значимости α. В противном случае гипотеза об отсутствии гетероскедастичности принимается, т.е. считается, что модель гомоскедастична.

Пример 5.1. Проверить гомоскедастичность/гетероскедастичность регрессионной модели, построенной по данным примера 2.2, на уровне значимости 0,01.

Решение. Для нахождения коэффициента ранговой корреляции Спирмена $r_{x,e}$ воспользуемся данными табл. 2.3 (см. пример 2.3), дополнив их необходимыми вычислениями:

Т а б л и ц а 5.1

i	x_i	e_i	$x_{(i)}$	$e_{(i)}$	r_{x_i}	r_{e_i}	d_i	d_i^2
1	107	–1,63	107	–4,45	1	2	–1	1
2	109	–0,49	109	–1,63	2	5	–3	9
3	110	1,57	110	–0,77	3	10	–7	49
4	113	0,77	113	–0,63	4	8	–4	16
5	120	–0,77	120	–0,49	5	3	2	4
6	122	–0,63	122	**0,22**	**6**	**4**	2	4
7	123	0,43	123	0,43	7	7	0	0
8	128	1,76	128	0,77	8	11	–3	9
9	136	1,29	136	1,29	9	9	0	0
10	140	–4,45	140	1,57	10	1	9	81
11	145	1,89	145	1,76	11	12	–1	1
12	150	0,22	150	1,89	12	6	6	36
Σ	–	–	–	–	–	–	–	210

По формуле (5.1) рассчитываем коэффициент ранговой корреляции Спирмена:

$$r_{x,e} = 1 - 6\cdot\frac{210}{12\cdot(12^2-1)} = 0{,}266,$$

а по формуле (5.2) – значение t-статистики:

$$t_r = \frac{0{,}266\sqrt{12-2}}{\sqrt{1-0{,}266^2}} = 0{,}873.$$

По табл.1 Приложения при $\alpha = 0{,}01$ и $n = 12$ находим $t_{1-\alpha/2}(n-2) = t_{0,995}(10) = 3{,}169$. Так как $|t_r| < t_{1-\alpha/2}(n-2)$ (0,873 < 3,169), то принимается гипотеза об отсутствии гетероскедастичности, т.е. на уровне значимости 0,01 можно считать дисперсии возмущений постоянными. ■

5.1.3. УСТРАНЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

При наличии гетероскедастичности для ее устранения необходимо преобразовать модель. Вид преобразования зависит от того, известны или нет дисперсии σ_i^2 случайных возмущений ε_i, $i = 1, 2, \ldots, n$.

Предположим, что общая линейная регрессионная модель

$$y_i = a_0 + \sum_{j=1}^{m} a_j x_{ji} + \varepsilon_i,\ i = 1,2,\ldots,n, \tag{5.3}$$

гетероскедастична, т.е. дисперсии σ_i^2 случайных возмущений не равны между собой, а сами возмущения ε_i и ε_j не коррелированны. Это означает, что ковариационная матрица (см. п. 3.2, тема 3) $\boldsymbol{K(\varepsilon)}$ вектора возмущений $\boldsymbol{\varepsilon}$ диагональная:

$$\boldsymbol{K}(\boldsymbol{\varepsilon})=\begin{pmatrix} \sigma_1^2 & 0 & ... & 0 \\ & \sigma_2^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sigma_n^2 \end{pmatrix}. \tag{5.4}$$

Если дисперсии возмущений σ_i^2 (i = 1, 2, …, n) известны, то гетероскедастичность устраняется достаточно легко. Действительно, будем рассматривать в качестве i-го наблюдения зависимой Y и объясняющих переменных X_j (j = 1, …, m) нормированные по σ_i значения этих переменных:

$$\tilde{y}_i = y_i/\sigma_i,\ \tilde{x}_{ji} = x_{ji}/\sigma_i,\ i=1,2,...,n.$$

Тогда модель (5.3) примет вид:

$$\tilde{y}_i=\tilde{a}_0+\sum_{j=1}^{m} a_j \tilde{x}_{ji}+\tilde{\varepsilon}_i,\ i=1,2,...,n, \tag{5.5}$$

где $\tilde{a}_0 = a_0/\sigma_i, \tilde{\varepsilon}_i = \varepsilon_i/\sigma_i$.

Очевидно, дисперсии $D(\tilde{\varepsilon}_i)=D(\varepsilon_i/\sigma_i)=D(\varepsilon_i)/\sigma_i^2=1,\ i=1,2,...,n$, т.е. модель (5.5) гомоскедастична. При таком преобразовании переменных ковариационная матрица $\boldsymbol{K(\varepsilon)}$ становится единичной, а сама модель (5.5) – классической.

Применение МНК к линейной регрессионной модели (5.5) дает эффективную оценку $\boldsymbol{a^*}$ вектора параметров $\boldsymbol{a}$ модели (5.3):

$$\boldsymbol{a}^* = \left(\boldsymbol{X}^T \boldsymbol{K}^{-1}(\boldsymbol{\varepsilon})\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{K}(\boldsymbol{\varepsilon})\boldsymbol{Y}, \tag{5.6}$$

где матрица $\boldsymbol{X}$ и вектор $\boldsymbol{Y}$ определены в п. 3.2 (тема 3).

Обобщение МНК для модели с гетероскедастичностью, когда ковариационная матрица диагональная, называется *взвешенным* (или *обобщенным*) *методом наименьших квадратов.*

На практике, однако, значения стандартных отклонений σ_i случайных возмущений модели почти никогда не бывают известными. В этом случае для применения взвешенного МНК, т.е. использования формулы (5.6), значения σ_i следует заменить их состоятельными оценками σ_i^*.

5.2. АВТОКОРРЕЛЯЦИЯ

5.2.1. СУЩНОСТЬ И ПОСЛЕДСТВИЯ АВТОКОРРЕЛЯЦИИ

Автокорреляция – это корреляционная зависимость между текущими значениями некоторой переменной и значениями этой же переменной, сдвинутыми на несколько периодов времени назад. Автокорреляция случайной составляющей ε модели – это корреляционная зависимость текущих ε_i и предыдущих ε_{i-l} значений случайной составляющей модели. Величина l называется *запаздыванием*, *сдвигом во времени* или *лагом*.

Автокорреляция случайных возмущений модели нарушает одну из предпосылок регрессионного анализа: условие

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j)=0,\ i,j=1,2,...,\ n;\ i\neq j,$$

не выполняется.

Автокорреляция может быть вызвана несколькими причинами, имеющими различную природу. Во-первых, иногда она связана с исходными данными и вызвана наличием ошибок измерения в значениях результирующей переменной. Во-вторых, в ряде случаев причину автокорреляции следует искать в формулировке модели. Модель может не включать фактор, оказывающий существенное воздействие на результат, влияние которого отражается на возмущениях, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t: автокорреляция обычно встречается при анализе временных рядов.

Постоянная направленность воздействия не включенных в модель переменных является наиболее частой причиной так называемой *положительной автокорреляции*.

Иллюстрацией положительной автокорреляции может служить следующий пример.

Пример 5.2. Пусть исследуется спрос Y на прохладительные напитки в зависимости от дохода X по ежемесячным и сезонным наблюдениям. Зависимость, отражающая увеличение спроса с ростом дохода, может быть представлена линейной функцией регрессии $y = ax + b$, изображенной вместе с результатами наблюдений на рис. 5.2.

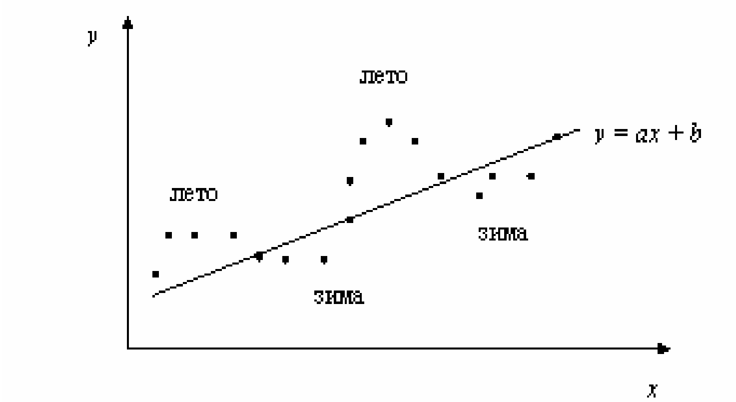


Рис. 5.2. Положительная автокорреляция

На величину спроса Y оказывают влияние не только доход X (учтенный фактор), но и другие факторы, которые не учтены в модели. Одним из таких факторов является время года.

Положительная автокорреляция означает постоянное в одном направлении действие неучтенных факторов на результирующую переменную. Так спрос на прохладительные напитки всегда выше линии регрессии летом (т.е. для летних наблюдений $\varepsilon > 0$) и ниже зимой (т.е. для зимних наблюдений $\varepsilon < 0$) (рис. 5.2). ■

Аналогичная картина может иметь место в макроэкономическом анализе с учетом циклов деловой активности.

Отрицательная автокорреляция означает разнонаправленное действие неучтенных в модели факторов на результат: за положительными значениями случайной составляющей ε в одних наблюдениях следуют, как правило, отрицательные в следующих, и наоборот. Графически это выражается в том, что результаты наблюдений y_i «слишком часто» «перескакивают» через график уравнения регрессии. Возможная схема рассеяния наблюдений в этом случае представлена на рис. 5.3.

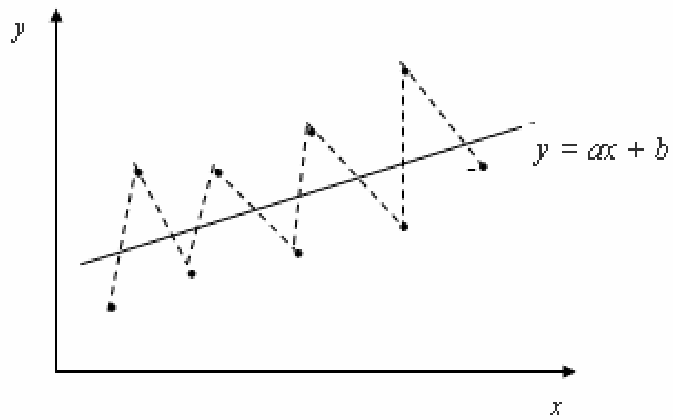


Рис. 5.3. Отрицательная автокорреляция

Последствия **автокорреляции в определенной степени сходны с последствиями гетероскедастичности. Среди них при применении МНК обычно выделяют следующие.**

1. МНК-оценки параметров, оставаясь несмещенными и линейными, перестают быть эффективными. Следовательно, они перестают обладать свойствами наилучших линейных несмещенных оценок.

2. Стандартные ошибки коэффициентов регрессии будут рассчитываться со смещением. Часто они являются заниженными, что влечет за собой увеличение *t*-статистик. Это может привести к признанию статистически значимыми объясняющих переменных, которые в действительности таковыми не являются. Смещенность возникает вследствие того, что выборочная остаточная дисперсия $S^2 = \sum e_i^2/(n-m-1)$ (*m* – число объясняющих переменных модели), которая используется при вычислении указанных величин (см. формулы (2.18) и (2.19)), является смещенной. Во многих случаях она занижает истинное значение дисперсии возмущений σ^2.

Вследствие вышесказанного все выводы, получаемые на основе соответствующих *t*- и *F*-статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получаемые при проверке качества оценок (параметров модели и самой модели в целом), могут быть ошибочными и приводить к неверным заключениям по построенной модели.

5.2.2. ОБНАРУЖЕНИЕ АВТОКОРРЕЛЯЦИИ

Если в модели присутствует автокорреляция, то наибольшее влияние на последующее наблюдение оказывает результат предыдущего наблюдения. Так, например, если рассматривается ряд значений курса какой-либо ценной бумаги, то именно результат последних торгов служит отправной точкой для формирования курса на следующих торгах.

Таким образом, отсутствие корреляции между соседними членами служит хорошим основанием считать, что корреляция отсутствует и в целом. Наиболее известным критерием (тестом) обнаружения автокорреляции между соседними членами является *критерий Дарбина – Уотсона*.

Критерий Дарбина–Уотсона основан на следующей идее: если корреляция случайных возмущений ε_i не равна нулю, то она присутствует и в остатках $e_i = y_i - y_i^*$, $i = 1, 2, \dots, n$. В критерии Дарбина – Уотсона для оценки корреляции используется *статистика Дарбина – Уотсона* вида

$$d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}\ . \qquad (5.7)$$

Статистика Дарбина–Уотсона *d* связана с выборочным коэффициентом корреляции $r = r_{e_i, e_{i-1}}$ между соседними значениями соотношением:

$$d \approx 2\,(1 - r).$$

Поскольку коэффициент корреляции *r* принимает значения $-1 \le r \le 1$, то $0 \le d \le 4$, и ее значения могут указывать на наличие либо отсутствие автокорреляции.

Действительно, если

- $r \approx 0$ (автокорреляция отсутствует), то $d \approx 2$;
- $r \approx 1$ (положительная автокорреляция), то $d \approx 0$;
- $r \approx -1$ (отрицательная автокорреляция), то $d \approx 4$.

Таким образом, необходимым условием отсутствия автокорреляции случайных возмущений является близость к 2 значения *d*-статистики Дарбина–Уотсона. Для ответа на вопрос, какие значения *d* можно считать статистически близкими к 2, разработаны специальные таблицы *нижних и верхних критических точек* распределения *d*-статистики Дарбина–Уотсона – $d_н = d_н(n, m, \alpha)$ и $d_в = d_в(n, m, \alpha)$ (см. табл.4 Приложения). Критические точки определяются в зависимости от количества наблюдений *n*, числа объясняющих переменных *m* и уровня значимости критерия α.

Алгоритм выявления автокорреляции возмущений на основе статистики Дарбина – Уотсона следующий.

1. Выдвигается гипотеза H_0 об отсутствии автокорреляции возмущений. Альтернативные гипотезы H_1 и H_1^* состоят соответственно в наличии положительной и отрицательной автокорреляции. Задается уровень значимости критерия α.

2. С помощью выборочного уравнения регрессии определяются значения остатков $e_i = y_i - y_i^*$ для каждого наблюдения $i = 1, 2, \dots, n$.

3. По формуле (5.7) рассчитывается значение d-статистики.

4. По таблице критических точек распределения d-статистики (см. Приложение) определяются критические точки $d_н = d_н(n, m, \alpha)$ и $d_в = d_в(n, m, \alpha)$ для заданного числа наблюдений n, числа объясняющих переменных m и уровня значимости α.

5. Делаются выводы по следующему правилу:

• если $0 \le d < d_н$, то принимается альтернативная гипотеза H_1 о положительной автокорреляции;

• если $d_н \le d < d_в$, то вопрос о принятии или отклонении основной гипотезы H_0 остается открытым (область неопределенности критерия);

• если $d_в \le d < 4 - d_в$, то принимается основная гипотеза H_0 об отсутствии автокорреляции;

• если $4 - d_в \le d < 4 - d_н$, то вопрос о принятии или отклонении основной гипотезы H_0 остается открытым (область неопределенности критерия);

• если $4 - d_н \le d \le 4$, то принимается альтернативная гипотеза H_1^* об отрицательной автокорреляции.

Изобразим данное правило графически:

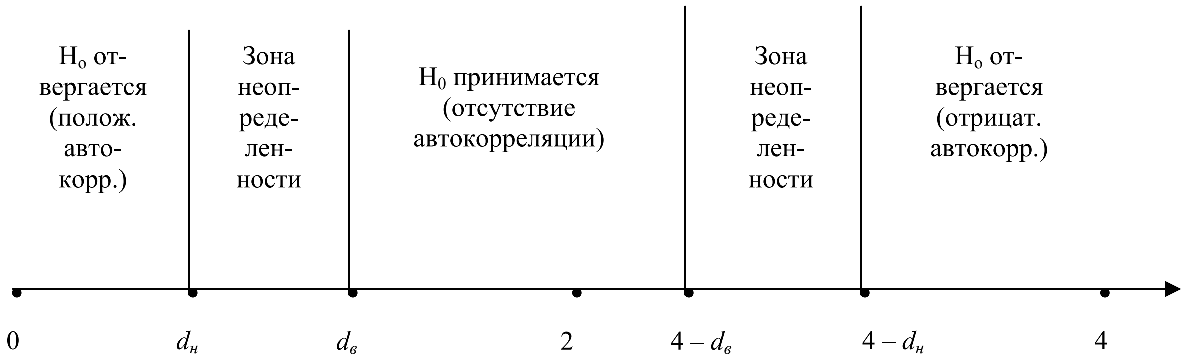


Рис. 5.4. Критерий Дарбина – Уотсона на автокорреляцию

Пример 5.3. Проверить наличие или отсутствие автокорреляции возмущений в регрессионной модели, построенной по данным примера 2.2, на уровне значимости 0,01.

Решение. Будем придерживаться вышеизложенной схемы критерия Дарбина – Уотсона.

1. Пусть основная гипотеза H_0 состоит в отсутствии автокорреляции возмущений, альтернативные H_1 и H_1^* – в наличии соответственно положительной и отрицательной автокорреляции. Зафиксируем уровень значимости критерия $\alpha = 0,01$.

2, 3. Воспользуемся данными табл. 2.3 примера 2.3 (или табл. 5.1 примера 5.1), дополнив их необходимыми вычислениями:

Таблица 5.2

i	e_i	e_{i-1}	$e_i - e_{i-1}$	$(e_i - e_{i-1})^2$	e_i^2
1	–1,63	–	–	–	2,66
2	–0,49	–0,49	**–1,14**	**1,30**	0,24
3	1,57	1,57	**–2,06**	**4,24**	2,46
4	0,77	0,77	**0,80**	**0,64**	0,59
5	–0,77	–0,77	1,54	2,37	0,59
6	–0,63	–0,63	–0,14	0,02	0,40
7	0,43	0,43	–1,06	1,12	0,18
8	1,76	1,76	–1,33	1,77	3,10
9	1,29	1,29	0,47	0,22	1,66
10	–4,45	–4,45	5,74	32,95	19,8
11	1,89	1,89	–6,34	40,20	3,57

12	0,22	0,22	1,67	2,79	0,05
Σ	≈ 0	–1,63	1,63	87,62	35,3

Вычисляем значение *d*-статистики по формуле (5.7):

$$d = \frac{87{,}62}{35{,}3} = 2{,}48\,.$$

4. По табл. 4 Приложения при $\alpha = 0{,}01$ и $n = 12$, $m = 1$ находим $d_н = 0{,}697$ и $d_в = 1{,}023$.

5. Так как $d_в \le d < 4 - d_в$ $(1{,}023 \le 2{,}48 < 2{,}977)$, то принимается основная гипотеза H_0 об отсутствии автокорреляции возмущений модели. ■

Устранение автокорреляции

Поскольку основная причина появления автокорреляции – это ошибки спецификации модели, то и методы устранения этого явления сводятся к устранению причин его возникновения. Во-первых, следует проанализировать, не упущен ли из рассмотрения какой-либо важный фактор или несколько факторов, влияющих на изменение зависимой переменной. Если такой фактор выявлен, следует рассмотреть новую модель с учетом данного фактора, найти оценки неизвестных параметров модели и вновь проверить наличие или отсутствие автокорреляции. Во-вторых, можно попробовать изменить форму модели. Например, вместо линейной рассмотреть квадратическую или логарифмическую. При этом новая форма модели может быть исследована как с первоначальным набором независимых переменных (объясняющих факторов), так и с новым, расширенным, набором факторов.

Если все разумные процедуры изменения спецификации модели, на ваш взгляд, исчерпаны, а автокорреляция имеет место, то можно предположить, что она обусловлена какими-то внутренними свойствами случайных возмущений. В этом случае можно воспользоваться так называемым авторегрессионным преобразованием, изложение которого, однако, выходит за рамки данного пособия.

5.3. МУЛЬТИКОЛЛИНЕАРНОСТЬ

5.3.1. СУТЬ И ПОСЛЕДСТВИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

Одним из условий множественного регрессионного анализа является предположение о линейной независимости объясняющих переменных модели, что означает линейную независимость столбцов матрицы регрессоров $\boldsymbol{X}$ (определение см. в п. 3.1). При нарушении данного условия, т.е. когда по крайней мере один из векторов-столбцов матрицы X есть линейная комбинация остальных, матрица $\boldsymbol{X}^T\boldsymbol{X}$ оказывается вырожденной, так как ее определитель равен нулю. Говорят, что в этом случае имеет место *полная коллинеарность* объясняющих переменных.

Наличие полной коллинеарности означает, что не существует обратная матрица $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, участвующая в основных соотношениях МНК, что приводит к невозможности решения соответствующей системы нормальных уравнений и получения оценок параметров регрессионной модели.

В практике экономических исследований полная коллинеарность встречается достаточно редко, так как ее несложно избежать уже на предварительной стадии анализа и отбора объясняющих переменных. Гораздо чаще приходится сталкиваться с ситуацией, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь. Матрица $\boldsymbol{X}^T\boldsymbol{X}$ в этом случае является невырожденной, но ее определитель очень мал. Тогда говорят о наличии *мультиколлинеарности* объясняющих переменных.

Мультиколлинеарность может привести к следующим нежелательным последствиям.

1. Вектор оценок $\boldsymbol{a}^*$ и его ковариационная матрица $\boldsymbol{K}(\boldsymbol{a}^*)$ в соответствии с формулами (3.10) и (3.12) пропорциональны обратной матрице $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, а значит, их элементы обратно пропорциональны величине определителя $det\left(\boldsymbol{X}^T\boldsymbol{X}\right)$. В результате получаются большие значения средних квадратических отклонений (стандартных ошибок) выборочных коэффициентов регрессии $a_0^*, a_1^*, \ldots,\ a_m^*$ и, следовательно, малые значения соответствующих *t*-статистик. Уменьшение *t*-статистик коэффициентов регрессии может привести к неоправданному выводу о существенности влияния соответствующей объясняющей переменной на зависимую переменную. Возможна ситуация, когда

оценка значимости коэффициентов регрессии по t-критерию не имеет смысла, хотя в целом регрессионная модель может оказаться значимой по F-критерию.

2. Оценки коэффициентов регрессии по МНК и их стандартные ошибки становятся очень чувствительными к незначительному изменению результатов наблюдений и объема выборки, т.е. они становятся *неустойчивыми*. Уравнения регрессии в этом случае, как правило, не имеют реального смысла, так как некоторые из его коэффициентов могут иметь неправильные с точки зрения теории знаки и неоправданно большие значения.

3. Становится затруднительным (а иногда и невозможным) определение влияния каждой из объясняющих переменных на результирующий показатель.

5.3.2. ОБНАРУЖЕНИЕ МУЛЬТИКОЛИНЕАРНОСТИ

Одним из простейших методов обнаружения мультиколлинеарности объясняющих переменных является вычисление и анализ выборочных коэффициентов парной корреляции между всеми парами объясняющих переменных: $r_{ij} = r_{x_i, x_j}$, $i, j = 1, 2, \ldots, m$ (см. формулу (2.33), п. 2.6, тема 2). Если мы расположим все коэффициенты парной корреляции в виде матрицы, то получим *матрицу корреляций* или *корреляционную матрицу*:

$$\boldsymbol{R} = \begin{pmatrix} 1 & r_{12} & r_{13} & \ldots & r_{1m} \\ r_{21} & 1 & r_{23} & \ldots & r_{2m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{m1} & r_{m2} & r_{m3} & \ldots & 1 \end{pmatrix}. \tag{5.8}$$

Поскольку $r_{jj} = r_{x_j, x_j} = 1$, $j = 1, 2, \ldots, m$, то на главной диагонали матрицы $\boldsymbol{R}$ расположены единицы. Кроме того, в матрице $\boldsymbol{R}$ элементы, симметричные относительно главной диагонали, равны между собой: $r_{ij} = r_{ji}$, $i, j = 1, 2, \ldots, m$.

Если для некоторых индексов i и j ($i \neq j$) значение $|r_{ij}|$ близко к единице, то имеются основания предполагать, что между переменными X_i и X_j существует сильная стохастическая зависимость, близкая к линейной, и, следовательно, имеет место мультиколлинеарность.

Данное правило достаточно обосновано для моделей с двумя объясняющими переменными X_1 и X_2. Если объясняющих переменных больше, то вывод о наличии мультиколлинеарности, основанный на близости к единице какого-то коэффициента парной корреляции, может быть ошибочным. Действительно, высокий положительный коэффициент парной корреляции между переменными X_i и X_j может быть не потому, что одна из них стимулирует изменение другой, а потому, что обе эти переменные изменяются в одном направлении под влиянием других переменных, как учтенных в модели, так и, возможно, неучтенных. Поэтому необходимо измерять действительную силу линейной связи между двумя переменными, очищенную от влияния на рассматриваемую пару переменных других факторов. Для измерения такой связи используют так называемый *частный коэффициент корреляции* (см. п. 3.5, тема 3).

Для оценки мультиколлинеарности объясняющих переменных можно использовать определитель $det\,\boldsymbol{R}$ корреляционной матрицы $\boldsymbol{R}$. Пусть основная гипотеза H_0 состоит в отсутствии мультиколлинеарности. Если выполнены основные предпосылки 1^0 – 6^0 множественного регрессионного анализа (см. п.3.2, тема 3), то доказано, что статистика

$$\chi^2 = -\left(n - 1 - \frac{2m+5}{6}\right)\ln(det\,\boldsymbol{R}) \tag{5.9}$$

при сравнительно небольшом числе наблюдений n имеет приближенно χ^2- распределение с $k = m(m - 1)/2$ степенями свободы. Следовательно, выполнение неравенства

$$\chi^2 \geq \chi^2_{1-\alpha}(k), \; k = m(m-1)/2,$$

означает наличие мультиколлинеарности объясняющих переменных на уровне значимости α.

Пример 5.4. Проверить наличие мультиколлинеарности между мощностью пласта X_1 (м) и уровнем механизации работ X_2 (%) при добыче угля в шахтах (по данным примера 3.1, тема 3) на уровне значимости 0,01.

Решение. Для расчета коэффициента корреляции между переменными X_1 и X_2 на основе табл. 3.1 составляем вспомогательную таблицу:

Таблица 5.3

i	x_{1i}	x_{2i}	x_{1i}^2	x_{2i}^2	$x_{1i} \cdot x_{2i}$
1	8	5	64	25	40
2	11	8	121	64	88
3	12	8	144	64	96
4	9	5	81	25	45
5	8	7	64	49	56
6	8	8	64	64	64
7	9	6	81	36	54
8	9	4	91	16	36
9	8	5	64	25	40
10	12	7	144	49	84
$\sum$	94	63	908	417	603

Вычислим вначале все необходимые средние:

$$\overline{x_1} = \frac{1}{10}\sum_{i=1}^{10} x_{1i} = \frac{94}{10} = 9{,}4;\ \overline{x_2} = \frac{1}{10}\sum_{i=1}^{10} x_{2i} = \frac{63}{10} = 6{,}3;$$

$$\overline{x_1^2} = \frac{1}{10}\sum_{i=1}^{10} x_{1i}^2 = \frac{908}{10} = 90{,}8;\ \overline{x_2^2} = \frac{1}{10}\sum_{i=1}^{10} x_{2i}^2 = \frac{417}{10} = 41{,}7;$$

$$\overline{x_1 \cdot x_2} = \frac{1}{10}\sum_{i=1}^{10} x_{1i} \cdot x_{2i} = \frac{603}{10} = 60{,}3\,;$$

затем дисперсии:

$$S_{x_1}^2 = \overline{x_1^2} - \left(\overline{x_1}\right)^2 = 90{,}8 - 9{,}4^2 = 2{,}44;\ S_{x_2}^2 = \overline{x_2^2} - \left(\overline{x_2}\right)^2 = 41{,}7 - 6{,}3^2 = 2{,}01$$

и средние квадратические отклонения:

$$S_{x_1} = \sqrt{2{,}44} = 1{,}56;\ S_{x_2} = \sqrt{2{,}01} = 1{,}42\,.$$

Наконец, по формуле (2.33) определяем

$$r_{12} = r_{x_1, x_2} = \frac{60{,}3 - 9{,}4 \cdot 6{,}3}{1{,}56 \cdot 1{,}42} = 0{,}4875\,.$$

Поскольку полученное значение коэффициента корреляции не высокое, то, возможно, мультиколлинеарности переменных X_1 и X_2 нет. Проверим эту гипотезу по критерию «хи-квадрат».

Корреляционная матрица $\boldsymbol{R}$ в данном случае имеет вид

$$\boldsymbol{R} = \begin{pmatrix} 1 & r_{12} \\ r_{21} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0{,}4875 \\ 0{,}4875 & 1 \end{pmatrix},$$

а ее определитель $det\,\boldsymbol{R} = 1 - 0{,}4875^2 = 0{,}238$.

По формуле (5.9) рассчитаем значение χ^2-статистики:

$$\chi^2 = -\left(10 - 1 - \frac{2 \cdot 2 + 5}{6}\right) \cdot \ln 0{,}238 = 10{,}77\,.$$

С помощью табл. 2 Приложения при α = 0,01 и $k = m(m-1)/2 = 2 \cdot 1/2 = 1$ находим $\chi^2_{1-\alpha}(k) = \chi^2_{0,99}(1) = 6{,}635$. Так как $\chi^2 > \chi^2_{1-\alpha}(k)$ (10,77 > 6,635), то гипотеза об отсутствии мультиколлинеарности отвергается на уровне значимости 0,01. ■

5.3.3. МЕТОДЫ УСТРАНЕНИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

Для устранения мультиколлинеарности разработаны разнообразные методы. Однако, прежде чем сформулировать некоторые из них, отметим, что в ряде случаев мультиколлинеарность не является настолько существенным недостатком исследуемой модели, чтобы прилагать значительные усилия по ее выявлению и устранению. В основном все зависит от целей исследования.

Если основная задача модели – прогноз будущих значений зависимой переменной, то при достаточно большом коэффициенте детерминации R^2 (обычно не менее 0,9) наличие мультиколлинеарности мало сказывается на прогнозных качествах модели.

Если же целью исследования является определение степени влияния каждой из объясняющих переменных на результирующую переменную, то наличие мультиколлинеарности, приводящее к увеличению стандартных ошибок, скорее всего, исказит истинные зависимости между переменными. В этой ситуации мультиколлинеарность является серьезной проблемой.

Отметим также, что не существует единого (универсального) метода устранения мультиколлинеарности, который был бы пригоден в любом случае. Это связано с тем, что причины мультиколлинеарности неоднозначны и во многом зависят от выборки.

а) исключение переменной (ых) из модели

Простейшим методом устранения мультиколлинеарности является исключение из модели одной или нескольких коррелированных переменных.

Однако необходима определенная осмотрительность при использовании данного метода, так как в этой ситуации возможны ошибки спецификации. Например, при исследовании спроса на некоторый товар (или услугу) в качестве объясняющих переменных можно использовать цену данного товара (услуги) и цены заменителей, которые зачастую коррелируют между собой. Исключив из модели цены заменителей, мы, скорее всего, допустим ошибку спецификации. Вследствие этого можно получить смещенные оценки и сделать ошибочные выводы. Поэтому в прикладных эконометрических моделях желательно не исключать объясняющие переменные до тех пор, пока мультиколлинеарность не станет серьезной проблемой.

б) получение дополнительных данных или новой выборки

Поскольку мультиколлинеарность напрямую зависит от выборки, то, возможно, при другой выборке мультиколлинеарности не будет либо она не будет столь серьезной.

Иногда для уменьшения мультиколлинеарности достаточно увеличить объем выборки. Увеличение количества данных уменьшает дисперсии коэффициентов регрессии и тем самым увеличивается их статистическая значимость.

Однако получение новой выборки или расширение старой не всегда возможно или связано с серьезными издержками.

в) изменение спецификации модели

В ряде случаев проблема мультиколлинеарности может быть решена изменением спецификации модели: либо изменяется форма модели, либо добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на результирующую переменную.

г) преобразование переменных

В ряде случаев проблему мультиколлинеарности можно устранить или уменьшить ее негативное влияние на модель с помощью преобразования переменных.

Пусть, например, в модели $Y = a_0 + a_1 X_1 + a_2 X_2 + \varepsilon$ переменные X_1 и X_2 сильно коррелируют между собой.

В этой ситуации можно рассмотреть регрессионные модели зависимости относительных величин:

$$Y/X_1 = a_0 + a_1 X_2 / X_1 + \varepsilon,\ Y / X_2 = a_0 + a_1 X_1 / X_2 + \varepsilon.$$

Вполне вероятно, что в этих моделях проблема мультиколлинеарности будет отсутствовать. Возможны и другие преобразования, близкие по своей сути к вышеописанным.

Контрольные вопросы

1. В чем суть гетероскедастичности?
2. Какое из следующих утверждений верно, ложно или не определено:

а) вследствие гетероскедастичности оценки перестают быть эффективными и состоятельными;

б) оценки и дисперсии оценок остаются несмещенными;

в) выводы по t- и F-статистикам являются ненадежными;

г) при наличии гетероскедастичности стандартные ошибки оценок будут заниженными;

д) гетероскедастичность проявляется через низкое значение d-статистики Дарбина – Уотсона;

е) тест ранговой корреляции Спирмена основан на использовании t-статистики;

ж) использование взвешенного метода наименьших квадратов (ВМНК) носит ограниченный характер, так как для его использования необходимо знать дисперсии возмущений.

3. В чем суть взвешенного метода наименьших квадратов (ВМНК)?
4. Объясните, почему при наличии гетероскедастичности ВМНК позволяет получить более эффективные оценки, чем обычный МНК.
5. Что такое автокорреляция? Назовите основные причины автокорреляции.
6. Что может вызвать положительную (отрицательную) автокорреляцию?
7. Какая предпосылка регрессионного анализа нарушается при автокорреляции?
8. Каковы последствия автокорреляции?
9. Опишите схему использования d-статистики Дарбина – Уотсона.
10. Верны или ложны следующие утверждения? Ответы поясните.

а) автокорреляция характерна в основном для временных рядов;

б) при наличии автокорреляции оценки, полученные по МНК, являются смещенными;

в) статистика Дарбина – Уотсона d лежит в пределах от 0 до 4;

г) при наличии автокорреляции значение статистики Дарбина – Уотсона d близко к 2.

11. Объясните значения понятий «полная коллинеарность» и «мультиколлинеарность».
12. Каковы основные последствия мультиколлинеарности?
13. Как можно обнаружить мультиколлинеарность?
14. Перечислите основные методы устранения мультиколлинеарности.
15. Какие из следующих утверждений верны, ложны или не определены? Ответ поясните.

а) при наличии высокой мультиколлинеарности невозможно оценить статистическую значимость коэффициентов регрессии при коррелированных переменных;

б) наличие мультиколлинеарности не является препятствием для получения по МНК эффективных оценок;

в) мультиколлинеарность не является существенной проблемой, если основная задача построенной регрессионной модели состоит в прогнозировании значений результирующей переменной;

г) высокие значения коэффициентов парной корреляции между объясняющими переменными не всегда являются признаком мультиколлинеарности;

д) при наличии мультиколлинеарности оценки коэффициентов регрессии остаются несмещенными, но их t-статистики будут слишком низкими;

е) пультиколлинеарность не приводит к получению смещенных коэффициентов регрессии, но ведет к получению смещенных оценок для дисперсий этих коэффициентов;

д) в регрессионной модели $Y = a_0 + a_1 X_1 + a_2 X_2 + \varepsilon$ наличие мультиколлинеарности можно обнаружить, если вычислить коэффициент корреляции между X_1 и X_2.

Контрольные задания

1. Проводится анализ зависимости средней заработной платы (з/п) Y от средней производительности X на предприятиях различного масштаба. Проведенное обследование нашло отражение в следующей таблице.

Количество сотрудников предприятия	Средняя Производительность (усл. ед.)	Средняя з/п (усл. ед.)	Стандартное отклонение з/п (усл. ед.)
1 ÷ 4	9 320	3 320	740
5 ÷ 9	8 630	3 640	850
10 ÷ 19	8 050	3 900	730
20 ÷ 49	8 320	4 120	820
50 ÷ 99	8 600	4 090	950
100 ÷ 199	9 120	4 200	1 100
200 ÷ 499	9 540	4 380	1 250
500 ÷ 999	9 730	4 500	1 290
1000 ÷ 1999	10 120	4 610	1 350
2000 ÷ 4999	10 740	4 800	1 100
Не менее 5000	11 200	5 000	1 520

а) оцените параметры модели парной линейной регрессии $Y = aX + b + \varepsilon$, используя обычный МНК;

б) оцените параметры преобразованной модели регрессии $Y/\sigma = a \cdot X/\sigma + \varepsilon/\sigma$, используя ВМНК;

в) сравните полученные результаты. Какую из моделей вы предпочтете и почему?

2. Пусть в модели парной линейной регрессии $Y = aX + b + \varepsilon$ имеет место соотношение $D(\varepsilon_i) = \sigma^2 \cdot x_i^2$, $i = 1, 2, \ldots, n$.

а) Будет ли иметь место в данной ситуации гетероскедастичность?

б) Какое преобразование можно предложить, чтобы устранить проблему гетероскедастичности?

в) Опишите поэтапно предложенную схему.

3. Известны данные по доходам X и Y (в усл. ед.) на непродовольственные товары для 12 домохозяйств:

X	26,2	33,1	42,5	47,0	48,5	49,0	49,1	50,9	52,4	53,2	54,0	58,0
Y	10,0	11,2	15,4	20,5	21,2	19,5	23,0	19,0	19,5	18,0	24,5	34,6

а) определите по МНК оценки параметров парной линейной модели регрессии Y на X;

б) оцените качество построенной модели;

в) проведите графический анализ остатков на наличие гетероскедастичности;

г) проверьте выводы, полученные в предыдущем пункте, с помощью критерия ранговой корреляции Спирмена на уровне значимости 0,05;

д) примените для указанных статистических данных ВМНК, предполагая, что $D(\varepsilon_i) = \sigma^2 \cdot x_i^2$, $i = 1, 2, \ldots, n$;

е) определите, существенно ли повлияла гетероскедастичность на качество модели, построенном по МНК.

4. Выдвигается предположение, что средняя заработная плата (з/п) наемных рабочих пропорциональна их стажу. Для анализа данного утверждения обследуются 20 рабочих восьми категорий стажа. Получены следующие статистические данные:

Стаж	[0,5)	[5,10)	[10,15)	[15,20)	[20,25)	[25,30)	[30,35)	[35,40]
З/п (тыс.руб)	10,0	12,5	14,3	18,7	25,4	29,0	32,0	34,3

а) постройте модель линейной регрессии, в которой заработная плата является зависимой, а стаж работы – объясняющей переменной (модель строится в предположении, что дисперсии возмущений постоянны);

б) оцените качество построенной модели;

в) есть ли основания считать, что для данной регрессионной модели весьма вероятна гетероскедастичность? Если да, то почему?

г) предполагая, что дисперсия возмущений пропорциональна трудовому стажу, постройте на основании тех же данных уравнение регрессии по ВМНК;

д) предполагая, что дисперсия возмущений пропорциональна квадрату величины трудового стажа, постройте по ВМНК соответствующее уравнение регрессии.

5. Пусть при 50 наблюдениях и 3 объясняющих переменных d-статистика Дарбина – Уотсона принимает следующие значения:

а) 0,91;

б) 1,37;

в) 2,34;

г) 3,01;

д) 3,72.

Не обращаясь к таблице критических точек распределения Дарбина – Уотсона, выскажите мнение о наличии автокорреляции. Проверьте свои выводы по таблице.

6. По статистическим данным за 20 лет построена линейная регрессионная модель зависимости между ценой бензина и объемом продаж бензина. С помощью этой модели вычислена d-статистика Дарбина – Уотсона, равная 0,71.

а) Будет ли в данном случае иметь место автокорреляция остатков? Если да, она положительная или отрицательная?

б) Что могло послужить причиной автокорреляции?

в) Какими будут ваши рекомендации по совершенствованию модели с целью устранения автокорреляции?

7. Приведены данные за 15 по темпам прироста заработной платы Y (%), производительности труда X_1 (%), а также по уровню инфляции X_2 (%):

Год	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_1	3,5	2,8	6,3	4,5	3,1	1,5	7,6	6,7	4,2	2,7	4,5	3,5	5,0	2,3	2,8
X_2	4,5	3,0	3,1	3,8	3,8	1,1	2,3	3,6	7,5	8,0	3,9	4,7	6,1	6,9	3,5
Y	9,0	6,0	8,9	9,0	7,1	3,2	6,5	9,1	14,6	11,9	9,2	8,8	12,0	12,5	5,7

а) оцените по МНК линейное уравнение регрессии Y на X_1 и X_2;

б) проверьте качество построенного уравнения регрессии с надежностью 0,95;

в) проведите проверку наличия в модели гетероскедастичности, автокорреляции и мультиколлинеарности на уровне значимости 0,05.

8. Имеется выборка из 10 наблюдений за переменными X_1 , X_2 , Y :

X_1	1	2	3	4	5	6	7	8	9	10
X_2	1	1,6	2,2	2,8	3,4	4	4,6	5,2	5,6	6,2
Y	0	3	6	9	12	15	18	21	24	27

а) можно ли по этим данным с помощью МНК оценить коэффициенты линейной регрессии с двумя объясняющими переменными? Ответ поясните;

б) в случае отрицательного ответа на предыдущий вопрос предложите преобразования, которые позволят оценить коэффициенты регрессии.

9. По выборке объема n = 50 для переменных X_1 , X_2 , X_3 получена следующая корреляционная матрица:

$$\boldsymbol{R}=\begin{pmatrix} 1{,}0 & 0{,}45 & -0{,}35 \\ 0{,}45 & 1{,}0 & 0{,}52 \\ -0{,}35 & 0{,}52 & 1{,}0 \end{pmatrix}.$$

а) найдите и оцените статистическую значимость частных коэффициентов корреляции $r_{12\bullet3}$, $r_{13\bullet2}$, $r_{23\bullet1}$;

б) при рассмотрении какой регрессии будет иметь место мультиколлинеарность?

Литература

1. *Доугерти К.* Введение в эконометрику. – М.: ИНФРА-М, 1999.–Гл. 5, 7.
2. *Воронович Н.В.*, *Русин Г.Л.* Эконометрика: Методические указания по выполнению контрольных работ. – Новосибирск, НГУЭУ, 2005.
3. *Практикум* по эконометрике /Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001. – II раздел.
4. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – Гл. 3, 6.

Тема 6. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

6.1. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

Многие сложные социально-экономические явления и процессы не всегда допускают адекватное описание с помощью только одного соотношения (уравнения). Кроме того, некоторые переменные могут оказаться настолько взаимосвязанными, что трудно однозначно определить, какая из них является зависимой, а какая независимой переменной. Поэтому при построении эконометрической модели в подобной ситуации используют системы уравнений. Выделяют следующие три вида эконометрических систем.

Система независимых уравнений, когда каждая зависимая переменная рассматривается как функция одного и того же набора независимых переменных:

$$\begin{cases} Y_1 = a_{10} + a_{11}X_1 + a_{12}X_2 + \ldots + a_{1m}X_m + \varepsilon_1, \\ Y_2 = a_{20} + a_{21}X_1 + a_{22}X_2 + \ldots + a_{2m}X_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_k = a_{k0} + a_{k1}X_1 + a_{k2}X_2 + \ldots + a_{km}X_m + \varepsilon_k. \end{cases}$$

Модель экономической эффективности сельскохозяйственного производства может служить примером такой модели. Здесь в качестве зависимых переменных выступают показатели, характеризующие эффективность сельскохозяйственного производства (урожайность с 1 га, продуктивность коров, себестоимость 1 ц молока и т.д.), а в качестве факторов – специализация хозяйства, количество голов на 100 га пашни, затраты труда и т. п.

Система рекурсивных уравнений, когда в каждом последующем уравнении системы зависимая переменная представляет функцию от всех зависимых и независимых переменных предшествующих уравнений:

$$\begin{cases} Y_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1m}X_m + \varepsilon_1, \\ Y_2 = b_{21}Y_1 + a_{21}X_1 + a_{22}X_2 + \ldots + a_{2m}X_m + \varepsilon_2, \\ Y_3 = b_{31}Y_1 + b_{32}Y_2 + a_{31}X_1 + a_{32}X_2 + \ldots + a_{3m}X_m + \varepsilon_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_k = b_{k1}Y_1 + a_{k2}Y_2 + \ldots + b_{k,k-1}Y_{k-1} + a_{k1}X_1 + a_{k2}X_2 + \ldots + a_{km}X_m + \varepsilon_k. \end{cases}$$

Примером такой системы может служить *модель производительности труда и фондоотдачи* вида:

$$\begin{cases} Y_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \varepsilon_1, \\ Y_2 = b_{21}Y_1 + a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \varepsilon_2, \end{cases}$$

где Y_1 – производительность труда; Y_2 – фондоотдача;
X_1 – фондовооруженность; X_2 – энерговооруженность;
X_3 – квалификация рабочих.

В приведенных двух видах систем каждое уравнение может рассматриваться самостоятельно, и параметры таких уравнений можно определить с помощью метода наименьших квадратов (МНК).

Система взаимозависимых (*совместных* или *одновременных*) *уравнений*, когда одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую часть системы:

$$\begin{cases} Y_1 = b_{12}Y_2 + a_{13}Y_3 + \ldots + b_{1k}Y_k + a_{11}X_1 + a_{12}X_2 + \ldots + a_{1m}X_m + \varepsilon_1, \\ Y_2 = b_{21}Y_1 + a_{23}Y_3 + \ldots + b_{2k}Y_k + a_{21}X_1 + a_{22}X_2 + \ldots + a_{2m}X_m + \varepsilon_2, \\ \ldots\ \ldots\ \ldots \\ Y_k = b_{k1}Y_1 + a_{k2}Y_2 + \ldots + b_{k,k-1}Y_{k-1} + a_{k1}X_1 + a_{k2}X_2 + \ldots + a_{km}X_m + \varepsilon_k. \end{cases}$$

Название «система одновременных уравнений» подчеркивает тот факт, что в системе одни и те же переменные *одновременно* рассматриваются как зависимые в одних уравнениях и как независимые в других. Такие системы наиболее распространены в эконометрических исследованиях.

В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения оценок его параметров традиционный МНК неприменим, так как нарушаются его предпосылки. С этой целью используются специальные приемы оценивания.

***Пример 6.1.* (*Модель Кейнса формирования доходов*).**

В простейшей макроэкономической модели Кейнса формирования национального дохода национальная экономика рассматривается как замкнутая система без государственного вмешательства:

$$\begin{array}{rl} \text{Функция потребления:} & \begin{cases} C_t = aY_t + b + \varepsilon_t, & (6.1.1) \\ Y_t = C_t + I_t, & (6.1.2) \end{cases} \\ \text{Макроэкономическое тождество:} & \end{array}$$

Здесь C_t , Y_t , I_t – объем потребления, совокупный национальный доход и инвестиции соответственно в момент времени t, а ε_t – случайный член. ■

***Пример 6.2.* (*Модель спроса и предложения*).**

Одна из простейших систем одновременных уравнений используется при моделировании спроса – предложения в рыночной экономике. В этом случае одновременно в момент времени t формируются спрос Q_t^d и предложение Q_t^s товара в зависимости от его цены P_t . Предполагая спрос и предложение товара линейно зависящими от цены, а спрос, кроме того, зависящим также линейно от дохода потребителей I_t , получим следующую систему:

$$\begin{array}{rl} \text{Функция спроса:} & \begin{cases} Q_t^d = a_0 + a_1P_t + a_2I_t + \varepsilon_{1t}, & (6.2.1) \\ Q_t^s = b_0 + b_1P_t + \varepsilon_{2t}, & (6.2.2) \\ Q_t^d = Q_t^s. & (6.2.3) \end{cases} \\ \text{Функция предложения:} & \\ \text{Условие равновесия:} & \end{array}$$

Очевидно, наличие случайных отклонений в данной модели связано в первую очередь с отсутствием ряда важных объясняющих переменных: цен сопутствующих товаров, цены ресурсов, налогов, вкусов потребителей и т.д.

Первые два уравнения (6.2.1) и (6.2.2), если их рассматривать отдельно, представляются вполне обычными. Мы можем оценить параметры регрессии для каждого из этих уравнений. Но в таком случае остается открытым вопрос о равенстве спроса и предложения, т.е. может не выполняться равенство (6.2.3). Поэтому оценка параметров отдельных уравнений в такой ситуации теряет смысл. ■

6.2. СОСТАВЛЯЮЩИЕ СИСТЕМ УРАВНЕНИЙ. СТРУКТУРНАЯ И ПРИВЕДЕННАЯ ФОРМЫ МОДЕЛИ

При рассмотрении систем одновременных уравнений переменные делятся на два больших класса – эндогенные и экзогенные. *Эндогенные переменные* – это переменные, значения которых определяются внутри модели. *Экзогенные переменные* – это внешние по отношению к модели переменные; их значения определяются вне модели, и поэтому они считаются заданными.

Такая классификация позволяет определять эндогенные и экзогенные переменные. Например, в кейнсианской модели формирования доходов (пример 6.1) переменные C_t и Y_t формируют свои значения, подчиняясь уравнениям (6.1.1) и (6.1.2), т.е. внутри модели, и потому они эндогенные. Между тем переменная I_t считается в уравнении (6.1.2) заданной, ее значения формируются вне модели, она экзогенная. В модели спроса и предложения (пример 6.2) переменные Q_t^d и Q_t^s – эндогенные, переменная I_t – экзогенная.

С математической точки зрения, главное отличие между экзогенными и эндогенными переменными заключается в том, что *экзогенные переменные не коррелируют со случайными возмущениями модели*; между тем как эндогенные переменные могут коррелировать (и, как правило, коррелируют).

Уравнения, составляющие исходную модель, называют *структурными уравнениями модели*. Обычно их подразделяют на *поведенческие уравнения* и *уравнения-тождества*. В первых из них описываются взаимодействия между переменными, во вторых – соотношения, которые должны выполняться во всех случаях. Например, в модели (6.2) уравнения (6.2.1) и (6.2.2) – поведенческие, а (6.2.3) – тождество.

Структурная форма эконометрической модели описывает одно- и многосторонние статистические отношения между экономическими величинами в их непосредственном виде. Она содержит всю существенную информацию о зависимостях между экономическими явлениями и процессами.

Уравнения, в которых отражена схема определения эндогенных переменных, называются *уравнениями в приведенной форме* (*приведенными уравнениями*). Это уравнения, в которых эндогенные переменные выражены только через экзогенные или предопределенные переменные, а также случайные составляющие. *Предопределенными* называются экзогенные переменные и такие эндогенные переменные, значения которых определены до рассмотрения соотношения (например, из итерационных процедур).

Приведем общий вид системы одновременных уравнений в структурной форме. Пусть Y_1, Y_2, …, Y_k – эндогенные переменные, X_1, X_2, …, X_m – экзогенные переменные. Введем матрицы $\boldsymbol{B}$ и $\boldsymbol{A}$ вида:

$$\boldsymbol{B}=\begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \dots & \dots & \dots & \dots \\ b_{k1} & b_{k2} & \dots & b_{kk} \end{pmatrix}; \quad \boldsymbol{A}=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{km} \end{pmatrix}.$$

Тогда общий вид системы одновременных уравнений представляется в *структурной матричной форме* как

$$\boldsymbol{B}\cdot\vec{\boldsymbol{Y}}+\boldsymbol{A}\cdot\vec{\boldsymbol{X}}=\boldsymbol{\varepsilon}\,, \qquad (6.3)$$

где

$$\vec{\boldsymbol{Y}}=\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_k \end{pmatrix};\ \ \vec{\boldsymbol{X}}=\begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_m \end{pmatrix};\ \ \boldsymbol{\varepsilon}=\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_k \end{pmatrix}.$$

Если матрица $\boldsymbol{B}$ невырожденная, т.е. $det\,\boldsymbol{B} \neq 0$, то существует обратная матрица $\boldsymbol{B}^{-1}$. Благодаря этому можно решить систему уравнений (6.3) относительно зависимых переменных, умножая (6.3) слева на $\boldsymbol{B}^{-1}$:

$$\vec{\boldsymbol{Y}}+\boldsymbol{B}^{-1}\cdot\boldsymbol{A}\cdot\vec{\boldsymbol{X}}=\boldsymbol{B}^{-1}\cdot\varepsilon$$

или

$$\vec{\boldsymbol{Y}}=-\boldsymbol{B}^{-1}\cdot\boldsymbol{A}\cdot\vec{\boldsymbol{X}}+\boldsymbol{B}^{-1}\cdot\boldsymbol{\varepsilon}\,. \qquad (6.4)$$

Форма эконометрической модели, задаваемой в виде (6.4), является *матричной приведенной формой*. Если в (6.4) воспользоваться обозначениями

$$\boldsymbol{C}=-\boldsymbol{B}^{-1}\cdot\boldsymbol{A} \ \text{ и } \ \boldsymbol{u}=\boldsymbol{B}^{-1}\cdot\boldsymbol{\varepsilon}\,,$$

то приведенную матричную форму модели можно записать более просто:

$$\vec{\boldsymbol{Y}}=\boldsymbol{C}\cdot\vec{\boldsymbol{X}}+\boldsymbol{u}\,. \qquad (6.5)$$

Представим матричное уравнение (6.5) подробнее в виде системы уравнений:

$$\begin{cases} Y_1=c_{11}X_1+c_{12}X_2+\ldots+c_{1m}X_m+u_1\,, \\ Y_2=c_{21}X_1+c_{22}X_2+\ldots+c_{2m}X_m+u_2\,, \\ \quad\quad\ldots\ldots. \\ Y_k=c_{k1}X_1+c_{k2}X_2+\ldots+c_{km}X_m+u_k\,. \end{cases} \qquad (6.6)$$

Из (6.6) видно, что приведенная форма модели – это система независимых уравнений, в которой все эндогенные переменные модели выражены через предопределенные и возмущающие переменные модели. Поэтому параметры каждого из уравнений системы в приведенной форме можно определить обычным МНК. Приведенная форма строится для того, чтобы по МНК-оценкам ее параметров определить оценки структурных коэффициентов.

Пример 6.3. Представить кейнсианскую модель формирования доходов (пример 6.1) в приведенной форме.

Решение. Подставив C_t из (6.1.2) в (6.1.1), получим из структурной приведенную форму модели:

$$\begin{cases} Y_t=\lambda_0+\lambda_1 I_t+u_t\,, \\ C_t=\lambda_0+\lambda_2 I_t+u_t\,, \end{cases} \qquad (6.7)$$

где

$$\lambda_0=\frac{b}{1-a};\ \lambda_1=\frac{1}{1-a};\ \lambda_2=\frac{a}{1-a};\ u_t=\frac{\varepsilon_t}{1-a}.\ \blacksquare \qquad (6.8)$$

Пример 6.4. Представить структурные уравнения модели спроса и предложения (пример 6.2) в матричном виде и построить соответствующую приведенную форму.

Решение. В модели спроса и предложения можно положить $Q_t^s=Q_t^d=Q_t$. Введем следующие обозначения:

$$\vec{\boldsymbol{Y}}_t=\begin{pmatrix} Q_t \\ P_t \end{pmatrix};\ \ \vec{\boldsymbol{X}}_t=\begin{pmatrix} 1 \\ I_t \end{pmatrix};\ \ \boldsymbol{\varepsilon}_t=\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix};$$

$$\boldsymbol{B}=\begin{pmatrix} 1 & -a_1 \\ 1 & -b_1 \end{pmatrix};\ \ \boldsymbol{A}=\begin{pmatrix} -a_0 & -a_2 \\ -b_0 & 0 \end{pmatrix}.$$

Тогда в матричном виде структурная форма модели запишется в виде

$$\boldsymbol{B}\cdot\vec{\boldsymbol{Y}}_t+\boldsymbol{A}\cdot\vec{\boldsymbol{X}}_t=\boldsymbol{\varepsilon}_t .$$

Предположим, что $det\,\boldsymbol{B} = a_1 - b_1 \neq 0$, и определим матрицу

$$\boldsymbol{C}=-\boldsymbol{B}^{-1}\cdot\boldsymbol{A}=\frac{1}{b_1-a_1}\begin{pmatrix}-b_1 & a_1\\ -1 & 1\end{pmatrix}\cdot\begin{pmatrix}-a_0 & -a_2\\ -b_0 & 0\end{pmatrix}=\frac{1}{b_1-a_1}\begin{pmatrix}a_0b_1-a_1b_0 & a_2b_1\\ a_0-b_0 & a_2\end{pmatrix}.$$

По формуле (6.5) находим приведенную форму модели в матричном виде

$$\boldsymbol{Y}_t=\frac{1}{b_1-a_1}\begin{pmatrix}a_0b_1-a_1b_0 & a_2b_1\\ a_0-b_0 & a_2\end{pmatrix}\cdot\begin{pmatrix}1\\ I_t\end{pmatrix}=\frac{1}{b_1-a_1}\begin{pmatrix}a_0b_1-a_1b_0+a_2b_1I_t\\ a_0-b_0+a_2I_t\end{pmatrix}+\begin{pmatrix}u_{1t}\\ u_{2t}\end{pmatrix},$$

откуда

$$\begin{cases}Q_t=\lambda_0+\lambda_1 I_t+u_{1t},\\ P_t=\lambda_2+\lambda_3 I_t+u_{2t},\end{cases} \tag{6.9}$$

где

$$\lambda_0=\frac{a_0b_1-a_1b_0}{b_1-a_1};\ \lambda_1=\frac{a_2b_1}{b_1-a_1};\ \lambda_2=\frac{a_0-b_0}{b_1-a_1};\ \lambda_3=\frac{a_2}{b_1-a_1}; \tag{6.10}$$

$$u_{1t}=\frac{\varepsilon_{1t}-\varepsilon_{2t}}{b_1-a_1};\ u_{2t}=\frac{b_1\varepsilon_{1t}-a_1\varepsilon_{2t}}{b_1-a_1}. \blacksquare$$

6.3. ПРОБЛЕМА ИДЕНТИФИКАЦИИ. НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ ИДЕНТИФИЦИРУЕМОСТИ

Изменение формы уравнений модели позволяет устранить проблему коррелированности объясняющей переменной и случайного отклонения. Однако оно может привести к другой не менее серьезной проблеме – *проблеме идентификации*.

В п. 6.2 мы отмечали, что параметры приведенной формы могут оцениваться методом наименьших квадратов. Однако экономический смысл и интерес для анализа представляют **параметры структурной формы**. Именно структурная форма раскрывает экономический механизм формирования значений эндогенных переменных.

Структурный параметр называется *идентифицируемым*, если он может быть оценен однозначно (единственным способом) по параметрам приведенной модели. Уравнение *идентифицируемо*, если идентифицируемы все входящие в него структурные параметры; модель *идентифицируема*, если все ее уравнения идентифицируемы.

Структурный параметр называется *неидентифицируемым*, если его значение невозможно получить, даже зная точные значения параметров приведенной формы. Уравнение *неидентифицируемо*, если неидентифицируем хотя бы один входящий в него структурный параметр; модель *неидентифицируема*, если неидентифицируемо хотя бы одно ее уравнение.

Структурный параметр называется *сверхидентифицируемым*, если он не может быть оценен однозначно (единственным способом) по параметрам приведенной модели. Уравнение *сверхидентифицируемо*, если оценки его структурных параметров невозможно найти по параметрам приведенной модели; модель *сверхидентифицируема*, если среди уравнений модели есть хотя бы одно сверхидентифицируемое.

***Правила идентификации*(применяются только к структурной форме модели)**

Введем следующие обозначения:

K – число эндогенных переменных в модели;

k – число эндогенных переменных в данном уравнении;

M – число предопределенных переменных в модели;

m – число предопределенных переменных в данном уравнении;

$\boldsymbol{D}$ – матрица коэффициентов при переменных, не входящих в данное уравнение.

Необходимое (но не достаточное) условие идентифицируемости уравнения модели:

Для того чтобы уравнение модели было идентифицируемо, необходимо, чтобы число предопределенных переменных, не входящих в уравнение, было не меньше числа эндогенных переменных, входящих в уравнение, минус единица, т.е. $M - m \geq k - 1$.

Если $M - m = k - 1$, уравнение идентифицируемо.

Если $M - m > k - 1$, уравнение сверхидентифицируемо.

Достаточное условие идентифицируемости уравнения модели:

Для того чтобы уравнение было идентифицируемым, достаточно, чтобы ранг матрицы $\boldsymbol{D}$ *был равен* $(K - 1)$.

Напомним, что ранг матрицы – это размер наибольшей ее квадратной подматрицы, определитель которой не равен нулю, или, что эквивалентно, максимальное число линейно независимых строк (или столбцов) матрицы.

Сформулируем теперь *необходимые и достаточные* условия идентифицируемости уравнения модели:

1. Если $M - m > k - 1$ и ранг матрицы $\boldsymbol{D}$ равен $K - 1$, то уравнение сверхидентифицируемо.
2. Если $M - m = k - 1$ и ранг матрицы $\boldsymbol{D}$ равен $K - 1$, то уравнение идентифицируемо.
3. Если $M - m \geq k - 1$ и ранг матрицы $\boldsymbol{D}$ меньше $K - 1$, то уравнение неидентифицируемо.
4. Если $M - m < k - 1$, то уравнение неидентифицируемо. В этом случае ранг матрицы $\boldsymbol{D}$ меньше $K - 1$.

Пример 6.5. Проверить кейнсианскую модель формирования доходов (пример 6.1) на идентифицируемость.

Решение. В данной модели $K = 2$ (эндогенные переменные C_t и Y_t),

$M = 1$ (экзогенная переменная I_t).

Проверим выполнение необходимого условия идентификации.

Для 1-го уравнения (6.1.1) модели $k = 2$, $m = 0$. Поскольку $M - m = 1 - 0 = 1 = k - 1 = 2 - 1 = 1$, то уравнение идентифицируемо.

Достаточное условие идентификации также выполнено. Действительно, в первом уравнении (6.1.1) отсутствует только переменная I_t, поэтому матрица $\boldsymbol{D} = (1)$. Отсюда следует, что $rang(\boldsymbol{D}) = 1 = K - 1$ и уравнение идентифицируемо.

Второе уравнение (6.1.2) является тождеством и не подлежит идентификации, так как не содержит неизвестных параметров. Следовательно, модель Кейнса формирования доходов идентифицируема, т.е. ее параметры можно оценить однозначно с помощью приведенной формы модели (6.7). ■

Пример 6.6. Проверить модель спроса-предложения (пример 6.2) на идентифицируемость.

Решение. В данной модели $K = 2$ (эндогенные переменные Q_t и P_t),

$M = 1$ (экзогенная переменная I_t).

Проверим выполнение необходимого условия идентификации.

Для 1-го уравнения (6.2.1) модели $k = 2$, $m = 1$. Поскольку $M - m = 1 - 1 = 0 < k - 1 = 2 - 1 = 1$, то уравнение неидентифицируемо.

Для 2-го уравнения (6.2.2) модели $k = 2$, $m = 0$. Поскольку $M - m = 1 - 0 = 1 = k - 1 = 2 - 1 = 1$, то уравнение идентифицируемо. Нетрудно видеть, что для 2-го уравнения выполнено также достаточное условие идентифицируемости: $rang(\boldsymbol{D}) = 1 = K - 1$. ■

Замечания 6.1.

1. Идентифицируемость кейнсианской модели формирования доходов (пример 6.1) можно вывести из следующих соображений. Применяя МНК к системе приведенных уравнений (6.7), можно найти оценки их параметров λ_0, λ_1, λ_2 . Зная данные оценки, несложно определить оценки параметров a и b структурной модели (6.1), так как в силу (6.8) $a = \lambda_2/\lambda_1$, $b = \lambda_0/\lambda_1$.

2. Применим аналогичные рассуждения к модели спроса и предложения (6.2). Оценки параметров λ_0, λ_1, λ_2, λ_3 приведенной системы (6.9) можно получить с помощью МНК. Однако этого недостаточно для оценки пяти параметров a_0, a_1, a_2, b_0, b_1 системы структурных уравнений (6.2). Мы сможем определить параметры b_0 и b_1 функции предложения (6.2.2), используя соотношения (6.10): $b_1 = \lambda_1/\lambda_3$, $b_0 = \lambda_0 - b_1\lambda_2$. Однако a_0, a_1, a_2 определить однозначно невозможно. Для этого требуется некоторое доопределение модели.

3. Заметим, что проблема идентифицируемости – это проблема количества информации. В случае сверхидентифицируемости имеет место ситуация переопределенности, т.е. как бы «слишком много» информации, и противоречивость этой информации не позволяет получить искомое решение.

Между тем в ситуации неидентифицируемости информации «слишком мало», что дает возможность существованию нескольких различных решений. ◄

6.4. ОЦЕНКА СИСТЕМ УРАВНЕНИЙ

Непосредственное применение МНК для каждого из уравнений системы одновременных уравнений приводит, вообще говоря, к получению смещенных и несостоятельных оценок. Обычно это происходит вследствие коррелированности одной или нескольких объясняющих переменных со случайным отклонением.

В силу невозможности получения на основе «обычного» МНК качественных параметров системы одновременных уравнений необходимо использовать другие методы получения «хороших» оценок.

6.4.1. КОСВЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (КМНК)

Один из таких возможных методов – *косвенный метод наименьших квадратов (КМНК)*, основанный на использовании приведенных уравнений.

КМНК включает в себя следующие этапы:

1. Исходя из структурных уравнений, строятся уравнения в приведенной форме.
2. Оцениваются по МНК параметры уравнений в приведенной форме.
3. На основе оценок, найденных на этапе 2, оцениваются параметры структурных уравнений.

Пример 6.7. Для модели спроса и предложения (пример 6.2) на основании статистических данных, приведенных в табл.6.1, необходимо оценить коэффициенты функции предложения, используя для этого МНК и КМНК. (В табл. 6.1 p_t, q_t, i_t – значения переменных P, Q, I в момент времени t). Сравнить результаты.

Т а б л и ц а 6.1

t	1	2	3	4	5	Сумма	Среднее
p_t	1	2	3	4	5	15	3
q_t	8	10	7	5	1	31	6,2
i_t	2	4	3	5	2	16	3,2

Решение. В примере 6.6 показано, что функция предложения Q_t модели «спрос-предложение» (6.2) является идентифицируемой. В силу п. 2 замечаний 6.1 оценки b_1^* и b_0^* могут быть определены на основе оценок $\lambda_0^*, \lambda_1^*, \lambda_2^*, \lambda_3^*$:

$$b_1^* = \frac{\lambda_1^*}{\lambda_3^*}, \;\; b_0^* = \lambda_0^* - b_1^* \lambda_2^* .$$

Для наглядности вычислений построим вспомогательную табл. 6.2.

Т а б л и ц а 6.2

t	1	2	3	4	5	Сумма	Среднее
p_t	1	2	3	4	5	15	3
q_t	8	10	7	5	1	31	6,2
I_t	2	4	3	5	2	16	3,2
p_t^2	1	4	9	16	25	55	11

i_t^2	4	16	9	25	4	58	11,6
$p_t q_t$	8	20	21	20	5	74	14,8
$p_t i_t$	2	8	9	20	10	49	9,8
$q_t i_t$	16	40	21	25	2	104	20,8

По имеющимся статистическим данным оценим коэффициенты приведенных уравнений (6.9). Для этого воспользуемся формулами (2.7) и (2.8) парного регрессионного анализа (см. тему 2):

$$\lambda_1^* = \frac{\overline{i \cdot q} - \bar{i} \cdot \bar{q}}{\overline{i^2} - \bar{i}^2} = \frac{20{,}8 - 3{,}2 \cdot 6{,}2}{11{,}6 - 3{,}2^2} = \frac{0{,}96}{1{,}36} = 0{,}706;$$

$$\lambda_0^* = \bar{q} - \lambda_1^* \cdot \bar{i} = 6{,}2 - 0{,}706 \cdot 3{,}2 = 3{,}941;$$

$$\lambda_3^* = \frac{\overline{i \cdot p} - \bar{i} \cdot \bar{p}}{\overline{i^2} - \bar{i}^2} = \frac{9{,}8 - 3{,}2 \cdot 3}{11{,}6 - 3{,}2^2} = \frac{0{,}2}{1{,}36} = 0{,}147;$$

$$\lambda_2^* = \bar{p} - \lambda_3^* \cdot \bar{i} = 3 - 0{,}147 \cdot 3{,}2 = 2{,}529.$$

Следовательно, оценки параметров функции предложения по КМНК будут равны:

$$b_1^* = \frac{\lambda_1^*}{\lambda_3^*} = \frac{0{,}706}{0{,}147} = 4{,}803, \quad b_0^* = \lambda_0^* - b_1^* \lambda_2^* = 3{,}941 - 4{,}803 \cdot 2{,}529 = -8{,}206,$$

а оценка функции предложения будет иметь вид

$$q_t^* = -8{,}206 + 4{,}803 p_t.$$

В то же время, рассчитанное непосредственно по МНК оценки уравнения (6.2.2) будут:

$$b_1^* = \frac{\overline{p \cdot q} - \bar{p} \cdot \bar{q}}{\overline{p^2} - \bar{p}^2} = \frac{14{,}8 - 3 \cdot 6{,}2}{11 - 3^2} = \frac{-3{,}8}{2} = -1{,}9,$$

$$b_0^* = \bar{q} - b_1^* \bar{p} = 6{,}2 + 1{,}9 \cdot 3 = 11{,}9.$$

Тогда оценка функции предложения имеет вид:

$$q_t^* = 11{,}9 - 1{,}9 p_t.$$

Полученные результаты позволяют сделать вывод о том, что применение МНК в несоответствующих ситуациях может существенно исказить картину зависимости. ■

6.4.2. ДВУХШАГОВЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (ДМНК)

Если система одновременных уравнений сверхидентифицируема, то КМНК не используется, так как он не дает однозначных оценок параметров структурной модели. В этом случае могут использоваться другие методы оценивания, среди которых наиболее распространенным и простым является *двухшаговый метод наименьших квадратов* (*ДМНК*).

Основная идея ДМНК состоит в том, чтобы на основе приведенной формы модели получить для сверхидентифицируемого уравнения набор эндогенных переменных, а затем, используя их в структурной форме модели, обычным МНК получить оценки параметров. Метод получил название двухшагового, так как дважды используется МНК.

Алгоритм ДМНК включает следующие этапы:

Составление приведенной формы модели.

Применение обычного МНК к каждому уравнению приведенной формы и получение оценок приведенных параметров.

Определение теоретических значений эндогенных переменных, фигурирующих в качестве факторов в структурной форме модели.

Определение оценок структурных параметров каждого уравнения в отдельности обычным МНК, используя в качестве факторов входящие в это уравнение предопределенные переменные и теоретические значения эндогенных переменных, полученные на этапе 1.

Сверхидентифицируемая структурная модель может быть двух типов:

– все уравнения системы сверхидентифицируемые;

система содержит наряду со сверхидентифицируемыми также идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных параметров каждого уравнения используется ДМНК. Если в системе есть идентифицируемые уравнения, то структурные параметры по ним находятся из системы приведенных уравнений. ДМНК является наиболее общим методом решения систем одновременных уравнений. Для идентифицируемых уравнений ДМНК дает тот же результат, что и КМНК.

ДМНК обладает рядом достоинств, делающим его весьма привлекательным для практического применения.

1. В данном методе 2-й этап (этап оценки приведенных уравнений) применяется для конкретных уравнений, не затрагивая оставшиеся уравнения модели. Это позволяет уменьшать объем вычислений.

2. При наличии сверхидентифицируемых уравнений ДМНК в отличие от МНК определяет единственные оценки параметров модели.

3. Применяя данный метод, достаточно использовать лишь экзогенные и сверхидентифицируемые переменные модели.

Контрольные вопросы

1. Каковы основные причины использования систем уравнений для моделирования экономических явлений и процессов?
2. Назовите возможные способы построения систем уравнений. Чем они отличаются друг от друга?
3. Сформулируйте простейшие модели формирования доходов Кейнса и «спроса-предложения».
4. Как определяются эндогенные, экзогенные и предопределенныепеременные модели? Могут ли быть предопределенными:
   а) эндогенные;
   б) экзогенные переменные?
5. В чем состоит основное различие между структурными уравнениями системы и уравнениями в приведенной форме?
6. Напишите структурную и приведенную формы системы одновременных уравнений в матричном виде.
7. Напишите приведенные формы моделей формирования доходов Кейнса и «спроса-предложения».
8. Почему обычный МНК практически не используется для оценки систем одновременных уравнений?
9. В чем состоит проблема идентификации модели?
10. Приведите необходимые и достаточные условия идентифицируемости модели.
11. Назовите причины неидентифицируемости и сверхидентифицируемости систем структурных уравнений.
12. Являются ли идентифицируемыми модели: а) формирования доходов Кейнса; б) спроса и предложения?
13. Для оценки каких систем возможно использование обычного МНК?
14. В чем состоит суть косвенного метода наименьших квадратов (КМНК)?
15. В каких случаях используется двухшаговый метод наименьших квадратов (ДМНК)? Раскройте его содержание.
16. Какие из следующих утверждений являются истинными, ложными или не определенными? Ответ поясните.

а) обычный МНК неприменим для оценки коэффициентов структурных уравнений систем одновременных уравнений;

б) МНК редко используется для оценки коэффициентов структурных уравнений систем одновременных уравнений, так как существуют методы получения более качественных оценок;

в) экзогенные и предопределенные переменные модели по сути являются одним и тем же;

г) проблема неидентифицируемости в первую очередь связана с невозможностью получения оценок коэффициентов структурных уравнений;

д) если уравнения системы идентифицируемы, то оценки получаемые по КМНК и ДМНК, идентичны;

е) для идентифицируемых систем ДМНК не используется.

Контрольные задания

1. Рассматривается система уравнений вида

$$\begin{cases} Y_1 = aX + bY_2 + \varepsilon_1, \\ Y_2 = cY_1 + \varepsilon_2. \end{cases}$$

Проверить, является ли данная система идентифицируемой. Изменится ли ответ, если в число регрессоров второго уравнения включить: а) константу; б) переменную *X*.

2. К системе двух уравнений вида

$$\begin{cases} Y_1 = a_1 X_1 + b_1 Y_2 + \varepsilon_1, \\ Y_2 = a_2 X_2 + b_2 Y_1 + \varepsilon_2 \end{cases}$$

применен КМНК. Для коэффициентов приведенной системы

$$\begin{cases} Y_1 = c_1 X_1 + c_1 X_2 + v_1, \\ Y_2 = c_3 X_1 + c_4 X_1 + v_2 \end{cases}$$

получены следующие оценки: $c_1^* = 2{,}2$; $c_2^* = 0{,}4$; $c_3^* = 0{,}08$; $c_4^* = -0{,}5$.

Найти оценки коэффициентов исходной системы ДМНК.

3. Пусть макроэкономическая модель закрытой экономики представлена в следующем упрощенном виде:

$$\begin{cases} C_t = a_0 + a_1 Y_t + \varepsilon_t, \\ I_t = b_0 + b_1 R_t + v_t, \\ Y_t = C_t + I_t + G_t. \end{cases}$$

Здесь Y_t – ВНП в году *t*; C_t – объем потребления в году *t*; I_t – объем инвестиций в году *t*; G_t – объем государственных расходов в году *t*; R_t – процентная ставка в году *t*; ε_t и v_t – случайные составляющие модели.

а) Какие из указанных переменных данной модели являются экзогенными, а какие – эндогенными?

б) Является ли модель идентифицируемой?

в) Как можно оценить параметры модели?

4. Рассматривается модель «спрос-предложение» следующего вида:

$$\begin{array}{rl} \text{Спрос:} & \begin{cases} Q^d = a_0 + a_1 P + \varepsilon, \\ Q^s = b_0 + b_1 P + b_2 W + v, \\ Q^d = Q^s. \end{cases} \\ \text{Предложение:} & \\ \text{Условие равновесия:} & \end{array}$$

Здесь *Q* – количество товара;

P – цена товара;

W – заработная плата;

ε и v – случайные отклонения, удовлетворяющие предпосылкам регрессионного анализа.

Пусть имеются следующие наблюдения:

P	10	15	5	8	4

Q	6	6	18	12	8
W	2	6	2	7	4

а) Какие из переменных в данной модели являются экзогенными, а какие эндогенными?

б) Представьте данную систему в приведенном виде.

в) Определите по МНК коэффициенты приведенных уравнений (если это возможно).

г) Совпадают ли знаки найденных коэффициентов с предполагаемыми теоретически?

д) На основе найденных приведенных коэффициентов по КМНК определите структурные коэффициенты для функции спроса.

е) Можно ли по КМНК оценить структурные коэффициенты для функции предложения? Если да, то как?

Литература

1. *Доугерти К*. Введение в эконометрику. – М.: ИНФРА-М, 1999. – Гл. 11.
2. *Практикум* по эконометрике /Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001. – III раздел.
3. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – Гл. 4.

Тема 7. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Эконометрические модели можно построить, используя два типа исходных данных:

• данные, характеризующие совокупность различных объектов в определенный момент (период) времени;

• данные, характеризующие один объект за ряд последовательных моментов (периодов) времени, которые образуют *временной ряд*.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные на основе данных второго типа, называются *моделями временных рядов*.

Методы исследования пространственных моделей и моделей временных рядов, вообще говоря, существенно отличаются. Объясняется это следующим обстоятельством. В пространственных моделях выборка y_1, y_2, …, y_n значений результирующей переменной рассматривается как совокупность реализаций *одной* случайной величины Y. В моделях временных рядов выборка y_1, y_2, …, y_n представляет собой одну из реализаций (траекторий) случайного процесса $Y(t)$, т.е. это совокупность реализаций *многих* случайных величин $Y(t)$, зависящих от времени t. Следствием данного факта являются следующие принципиальные отличия временного ряда y_t (t = 1, 2, …, n) от последовательности наблюдений y_1, y_2, …, y_n, образующих случайную выборку:

• члены временного ряда, в отличие от элементов случайной выборки, как правило, не являются независимыми;

• члены временного ряда не являются одинаково распределенными.

7.1. ОСНОВНЫЕ ЭЛЕМЕНТЫ ВРЕМЕННОГО РЯДА И ЗАДАЧИ ИХ АНАЛИЗА

Под *временным рядом* (*динамическим рядом*, или *рядом динамики*) в экономике подразумевается совокупность наблюдений некоторого признака $Y(t)$, зависящего от времени t, в последовательные моменты времени t_1, t_2, …, t_n. Мы будем рассматривать временные ряды с *равноотстоящими* моментами наблюдений, т.е.

$$t_2 - t_1 = t_3 - t_2 = \ldots = t_n - t_{n-1} = \Delta,$$

где Δ – заданный временной такт (минута, час, сутки, неделя, месяц, квартал, год и т. д.). Поэтому в дальнейшем исследуемый временной ряд нам будет удобнее представлять в виде

$$y_1, y_2, \ldots, y_n, \tag{7.1}$$

где y_t – значение исследуемого признака, зарегистрированное в t-м такте времени (t = 1, 2, …, n), называемое также *уровнем* временного ряда.

При исследовании экономического временного ряда y_t в общем случае выделяются несколько его составляющих:

u_t – *тренд*, плавно меняющаяся компонента, описывающая влияние долговременных факторов, т.е. длительную (так называемую «вековую») тенденцию в изменении анализируемого признака $Y(t)$ (например, рост населения, экономическое развитие, изменение структуры потребления и т.п.);

v_t – *сезонная компонента*, отражающая повторяемость экономических процессов в течение не очень длительного периода (года, иногда месяца, недели и т. д.; например, объем продаж товаров или перевозок пассажиров в различные времена года);

c_t – *циклическая компонента*, отражающая повторяемость экономических процессов в течение длительных периодов (например, влияние демографических «ям», циклов солнечной активности и т.п.);

ε_t – *случайная компонента*, отражающая влияние не поддающихся учету и регистрации случайных факторов.

Отметим, что в отличие от ε_t первые три составляющие (компоненты) u_t, v_t, c_t являются *детерминированными*, *неслучайными*.

В большинстве случаев значение (или уровень) временного ряда y_t можно представить как сумму или произведение перечисленных выше компонент. Модель временного ряда первого типа называется *аддитивной*, второго – *мультипликативной*. Конечно, вовсе не обязательно, чтобы в процессе формирования значений всякого временного ряда участвовали одновременно компоненты всех четырех типов. Однако во всех случаях будет предполагаться непременное участие случайной компоненты ε_t. Кроме того, для определенности мы примем *аддитивную структуру* влияния компонент на формирование значений y_t, что означает правомерность представления значений членов временного ряда в виде разложения:

$$y_t = u_t + v_t + c_t + \varepsilon_t \ (t = 1, 2, ..., n)\,. \tag{7.2}$$

Выводы о том, участвуют или нет компоненты данного типа в формировании значений y_t, могут основываться как на анализе содержательной сути задачи, так и на специальном статистическом анализе исследуемого временного ряда.

Основные задачи анализа временного ряда состоят в том, чтобы по имеющейся траектории (7.1) анализируемого временного ряда y_t :

• определить, какие из неслучайных функций u_t, v_t и c_t присутствуют в разложении (7.2);

• построить «хорошие» оценки для тех неслучайных функций, которые присутствуют в разложении (7.2);

• подобрать модель, адекватно описывающую поведение случайных «остатков» ε_t, и статистически оценить параметры этой модели.

Отметим основные ***этапы*** анализа временных рядов:

графическое представление и описание поведения временного ряда;

выделение и удаление детерминированных (неслучайных) составляющих временного ряда (тренда, сезонных и циклических составляющих);

сглаживание и фильтрация (удаление низко – или высокочастотных составляющих временного ряда);

исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели для ее описания;

прогнозирование развития изучаемого процесса на основе имеющегося временного ряда.

7.2. ВЫЯВЛЕНИЕ СТРУКТУРЫ ВРЕМЕННОГО РЯДА

7.2.1. ПРОВЕРКА ГИПОТЕЗЫ О СУЩЕСТВОВАНИИ ТРЕНДА

Решение любой задачи по анализу временных рядов начинается с построения графика исследуемого показателя. При этом однако на графике не всегда четко прослеживается присутствие тренда во временном ряду. В таких случаях, прежде чем переходить к определению тенденции и выделению тренда, нужно выяснить, существует ли вообще тенденция в исследуемом процессе.

Основные подходы к решению этой задаче основаны на статистической проверке гипотез. Критерии выявления компонент ряда основаны на проверке гипотезы о *случайности* временного ряда, т.е. по существу на статистической проверке гипотезы

$$\mathrm{H}_0\text{: } M\,y_t = a = \text{const}. \tag{7.3}$$

Рассмотрим *критерий серий*, *основанный на медиане*, часто используемый на практике для проверки наличия/отсутствия тренда. Применение этого критерия может быть представлено в виде следующих шагов.

1. Из исходного ряда с уровнями $y_1, y_2, \ldots, y_n$ образуется ранжированный (упорядоченный по возрастанию) ряд $y_{(1)} \le y_{(2)} \le \ldots \le y_{(n)}$, где $y_{(1)}$ и $y_{(n)}$ – наименьшее и наибольшее значения соответственно из уровней исходного ряда $y_1, y_2, \ldots, y_n$.

2. Определяется медиана Me построенного ранжированного ряда. В случае нечетного значения длины ряда n ($n = 2m + 1$) медиана $Me = y_{(m+1)}$, в противном случае ($n = 2m$) медиана $Me = (y_{(m)} + y_{(m+1)})/2$.

3. Образуется последовательность δ_i из плюсов и минусов по следующему правилу:

$$\delta_i = \begin{cases} +, \text{ если } y_t > Me,\ t = 1,2,\ldots,n; \\ -, \text{ если } y_t < Me,\ t = 1,2,\ldots,n. \end{cases} \tag{7.4}$$

Если значение уровня исходного ряда y_t равно медиане Me, то это значение пропускается. Очевидно, общее число знаков «+» и «–» заранее неизвестно. Индекс i может принимать значения 1, 2, …, k, где $k \le n$.

4. Подсчитывается величина $\nu(n)$ – число серий в совокупности δ_i, где под серией понимается последовательность подряд идущих плюсов или минусов. Один плюс или один минус тоже считается серией.

Определяется $\tau_{\max}(n)$ – протяженность самой длинной серии.

5. Проверка гипотезы H_0 основывается на том факте, что при условии случайности временного ряда (при отсутствии тренда) протяженность самой длинной серии не должна быть слишком большой, а общее число серий – слишком маленьким. Поэтому для того чтобы не была отвергнута гипотеза о случайности исходного ременного ряда (об отсутствии тренда), должны выполняться следующие неравенства:

$$\begin{cases} \nu(n) > \left[\dfrac{1}{2}(n+1) - 1{,}96\sqrt{n-1} \right]; \\ \tau_{\max}(n) < [1{,}43\ln(n+1)], \end{cases} \tag{7.5}$$

где n – длина временного ряда;

$\nu(n)$ – число серий;

$\tau_{\max}(n)$ – протяженность самой длинной серии;

$[x]$ – целая часть числа x, т.е. целое число, ближайшее к x и не превосходящее его по модулю.

Если хотя бы одно из неравенств (7.5) нарушается, то гипотеза об отсутствии тренда временного ряда отвергается с вероятностью ошибки α, заключенной между 0,05 и 0,0975 и, следовательно, подтверждается наличие зависящей от времени неслучайной составляющей.

Пример 7.1. В следующей таблице приведены данные, отражающие спрос на некоторый товар (в усл. ед.) за восьмилетний период:

Т а б л и ц а 7.1

Год, t	1	2	3	4	5	6	7	8
Спрос, y_t	213	171	291	309	317	362	351	361

Изобразить графически временной ряд и проверить гипотезу о наличии тренда.

Решение. На рис.7.1 рассматриваемый временной ряд изображен графически в виде ломаной кривой. Из этого рисунка явная тенденция временного ряда не просматривается.

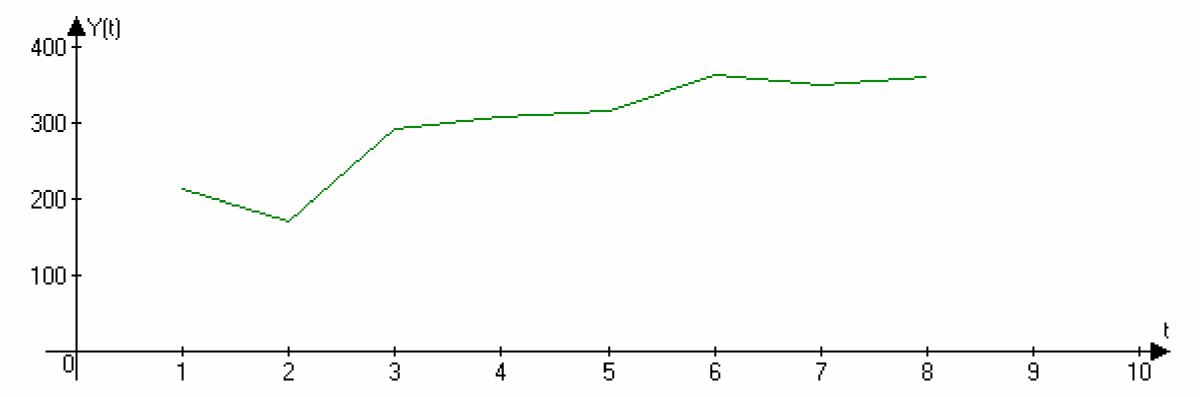


Рис.7.1. Временной ряд

Проверим нулевую гипотезу (7.3) об отсутствии тренда с помощью критерия серий, воспользовавшись описанной выше методикой.

1. Ранжировка уровней временного ряда представлена в табл.7.2:

Т а б л и ц а 7.2

i	1	2	3	4	5	6	7	8
$y_{(i)}$	171	213	291	309	317	351	361	362

2. Поскольку длина временного ряда n = 8 четная, то $m = n/2 = 4$, и медиана временного ряда $Me = (y_{(4)} + y_{(5)})/2 = (309 + 317)/2 = 313$.

3. Процесс формирования серий, в соответствии с (7.4), показан в следующей таблице:

Т а б л и ц а 7.3

i	1	2	3	4	5	6	7	8
δ_i	–	–	–	–	+	+	+	+

4. Анализ полученной последовательности знаков позволяет установить, что число серий $\nu(8) = 2$ и протяженность самой длинной серии $\tau_{\max}(8) = 4$.

5. Для проверки выполнения неравенств (7.5) определяем значения правых частей неравенств:

$$\left[\frac{1}{2}(8+1)-1{,}96\sqrt{8-1}\right]=[-0{,}686]=0;\quad [1{,}43\ln(8+1)]=[3{,}142]=3\,.$$

Тогда проверка выполнения условий (7.5) показывает, что первое неравенство выполняется, а второе – не выполняется. Следовательно, нулевая гипотеза (7.3) отвергается и делается вывод, что динамика временного ряда характеризуется наличием систематической составляющей – в изменении спроса присутствует тенденция.

Так как вероятность ошибочного вывода $0{,}05 \le \alpha \le 0{,}0975$, то надежность $\gamma = 1 - \alpha$ данного вывода заключена между 0,9025 и 0,95. ■

7.2.2. АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ

При наличии во временном ряде тренда и циклической компоненты значения каждого последующего уровня ряда становятся зависящими от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда*. Количественно ее можно измерить с помощью коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Пусть

$y_1, y_2, \ldots, y_{n-l}$ и $y_{1+l}, y_{2+l}, \ldots, y_n$ – две последовательности наблюдений временного ряда, сдвинутые друг относительно друга на l единиц, или, как говорят, с *лагом* l. Степень тесноты линейной связи между ними может быть оценена с помощью выборочного коэффициента корреляции

$$r(l)=\frac{(n-l)\sum_{t=1}^{n-l} y_t y_{t+l} - \sum_{t=1}^{n-l} y_t \sum_{t=1}^{n-l} y_{t+l}}{\sqrt{(n-l)\sum_{t=1}^{n-l} y_t^2 - \left(\sum_{t=1}^{n-l} y_t\right)^2}\sqrt{(n-l)\sum_{t=1}^{n-l} y_{t+l}^2 - \left(\sum_{t=1}^{n-l} y_{t+l}\right)^2}} . \qquad (7.6)$$

Так как коэффициент $r(l)$ измеряет корреляцию между членами одного и того же ряда, его называют *выборочным коэффициентом автокорреляции* (или просто *коэффициентом автокорреляции*), зависимость $r(l)$ – *автокорреляционной функцией*, а ее график – *коррелограммой*.

Лаг (сдвиг во времени) определяет *порядок* коэффициента автокорреляции. Если l = 1, то имеем коэффициент автокорреляции 1-го порядка; если l = 2, то коэффициент автокорреляции 2-го порядка и т. д.

При расчете $r(l)$ следует помнить, что с увеличением l число $(n - l)$ пар наблюдений y_t, y_{t+l} (t = 1, 2, …, $n - l$) уменьшается. Поэтому лаг l должен быть таким, чтобы число $(n - l)$ было достаточным для определения $r(l)$. Обычно ориентируются на соотношение $l \le n/4$.

Рассчитав несколько коэффициентов автокорреляции, можно определить лаг l, при котором автокорреляция $r(l)$ наиболее высокая, выявив тем самым *структуру временного ряда*. Если наиболее высоким оказывается $r(1)$, то исследуемый ряд содержит только тренд. Если наиболее высоким оказался $r(l)$, то ряд содержит (помимо тренда) колебания с периодом l. Если ни один из $r(1)$, $r(2)$, …, $r(l)$ не является значимым, можно сделать одно из двух предположений:

• либо временной ряд не содержит тренда и циклических колебаний, а его уровень определяется только случайной компонентой;

• либо ряд содержит сильный нелинейный тренд, для выявления которого нужен дополнительный анализ.

Пример 7.2. По данным временного ряда примера 7.1 выявить его структуру.

Решение. Поскольку $n/4$ = 8/4 = 2, то в соответствии с приведенной выше рекомендацией будем рассматривать коэффициенты автокорреляции $r(l)$ временного ряда для лагов l = 1, 2.

Найдем коэффициент автокорреляции 1-го порядка $r(1)$, т.е. коэффициент корреляции между последовательностями $(n - l)$ = 8 – 1 = 7 пар наблюдений y_t и y_{t+1} (t = 1, 2, …, 7):

y_t	213	171	291	309	317	362	351
y_{t+1}	171	291	309	317	362	351	361

Вычисляем необходимые суммы:

$$\sum_{t=1}^{7} y_t = 213+171+...+351=2\,014; \quad \sum_{t=1}^{7} y_t^2 = 213^2+171^2+...+351^2=609\,506;$$

$$\sum_{t=1}^{7} y_{t+1} = 171+291+...+361=2162; \quad \sum_{t=1}^{7} y_{t+1}^2 = 171^2+291^2+...+361^2=694\,458;$$

$$\sum_{t=1}^{7} y_t y_{t+1} = 213\cdot171+171\cdot291+...+351\cdot361=642\,583 .$$

Теперь по формуле (7.3) коэффициент автокорреляции 1-го порядка

$$r(1)=\frac{7\cdot642\,583-2014\cdot2162}{\sqrt{7\cdot609\,506-2014^2}\sqrt{7\cdot694\,458-2162^2}}=0,725 .$$

Коэффициент автокорреляции $r(2)$ 2-го порядка между членами ряда наблюдений y_t и y_{t+2} (t = 1, 2, …, 6) по шести парам наблюдений вычисляется аналогично: $r(2)$ = 0, 842.

Так как $r(2) > r(1)$, то полагаем, что анализируемый временной ряд содержит трендовую составляющую и циклическую составляющую с периодом 2 года. ■

7.3. СГЛАЖИВАНИЕ (ВЫРАВНИВАНИЕ) ВРЕМЕННОГО РЯДА

Одной из важнейших задач исследования экономического временного ряда является выявление *основной тенденции* изучаемого процесса, выраженной неслучайной составляющей $f(t)$ (тренда либо тренда с циклической или (и) сезонной компонентой). В некоторых случаях основная тенденция ясно прослеживается в динамике рассматриваемого показателя, в других ситуациях она может не просматриваться из-за ощутимых случайных колебаний. Например, сильные колебания в курсах акций в отдельные моменты времени могут заслонить наличие тенденции к росту или снижению этого показателя.

На практике для обнаружения основной тенденции часто используют простой прием – *укрупнение периода времени*. Например, ряд недельных данных можно преобразовать в ряд месячной динамики, ряд квартальных данных заменить годовыми уровнями. Уровни нового ряда могут быть получены суммированием уровней исходного ряда либо могут представлять средние значения.

При выявлении тенденции в развитии исследуемого явления используется распространенный прием – *сглаживание временного ряда*. Суть различных приемов сглаживания сводится к замене фактических уровней временного ряда расчетными, которые в меньшей степени подвержены колебаниям. Это способствует более четкому проявлению тенденции развития.

Методы сглаживания временных рядов можно условно разделить на два класса, опирающиеся на различные подходы: аналитический и алгоритмический.

7.3.1. АНАЛИТИЧЕСКОЕ ВЫРАВНИВАНИЕ ВРЕМЕННОГО РЯДА

В эконометрике большое внимание при выделении основной тенденции временного ряда уделяется методу *аналитического выравнивания* (*сглаживания*). Этот метод основан на допущении, что исследователь может задать общий вид функции $f(t)$, описывающей неслучайную составляющую временного ряда. В экономической практике такие функции часто называют *кривыми роста*. Их условно можно разделить на два класса в зависимости от того, какой тип динамики развития они отражают: *кривые роста без «насыщения» и кривые роста с «насыщением»*.

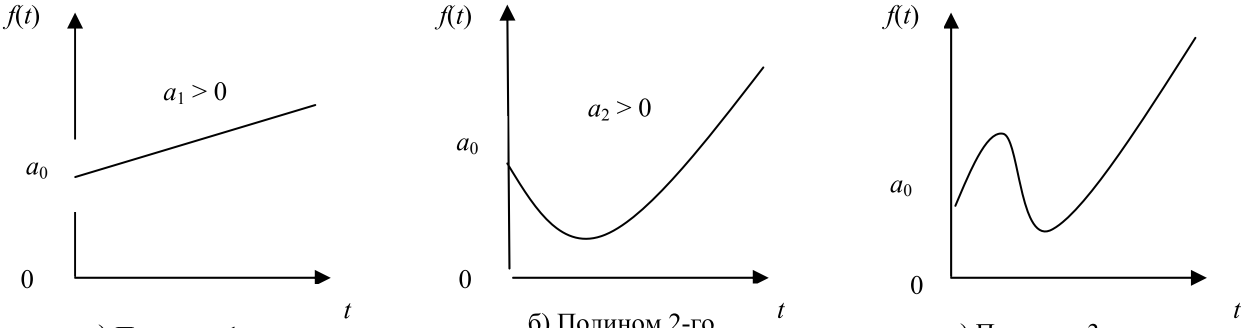
К первому классу относятся функции, используемые для описания процессов с монотонным (или кусочно-монотонным) характером тенденции развития и отсутствием пределов роста. Эти условия справедливы для многих экономических показателей, например, для большинства показателей промышленного производства в натуральном выражении. Среди кривых роста первого класса, прежде всего, следует выделить *полиномы* (*многочлены*):

$$f(t)=a_0+a_1t+a_2t^2+...+a_mt^m \ ,$$

где a_j ($j = 0, 1, \dots, m$) – коэффициенты и m – степень многочлена.

Коэффициенты полиномов невысоких степеней могут иметь конкретную интерпретацию в зависимости от содержания динамического ряда. Например, их можно трактовать как скорость роста (a_1), ускорение роста (a_2), изменение ускорения (a_3).

При выборе соответствующего полинома часто используют содержательный анализ (который может установить характер динамики процесса), а также визуальные наблюдения на основе графического изображения временного ряда (рис 7.2).



а) Полином 1-го порядка
$f(t) = a_0 + a_1 t$

б) Полином 2-го порядка
$f(t) = a_0 + a_1 t + a_2 t^2$

с) Полином 3-го порядка
$f(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$

Рис. 7.2. Полиномиальные кривые роста

Для подбора степени полинома может быть применен *метод последовательных разностей*. Суть этого метода состоит в вычислении разностей 1-го порядка $\Delta_t = y_t - y_{t-1}$, 2-го порядка $\Delta_t^{(2)} = \Delta_t - \Delta_{t-1}$, 3-го порядка $\Delta_t^{(3)} = \Delta_t^{(2)} - \Delta_{t-1}^{(2)}$ и т. д. Порядок разностей, при котором они будут примерно одинаковыми, принимается за степень полинома.

К представителям кривых роста первого класса относятся также показательные (экспоненциальные) функции, для которых характерна зависимость приростов от значений самой функции. Эти кривые хорошо описывают процессы, имеющие «лавинообразный» характер, когда прирост зависит от достигнутого уровня функции.

Простая показательная *(*экспоненциальная*) кривая имеет вид:*

$$f(t) = a \cdot b^t \quad (f(t) = e^{a_0 + a_1 \cdot t}).$$

Если $b > 1$ ($a_1 = \ln b > 0$), то кривая растет вместе с ростом t, и убывает, если $b < 1$ ($a_1 = \ln b < 0$) (рис. 7.3).

Параметр a ($a_0 = \ln a$) характеризует начальные условия развития, а параметр b – постоянный темп роста.

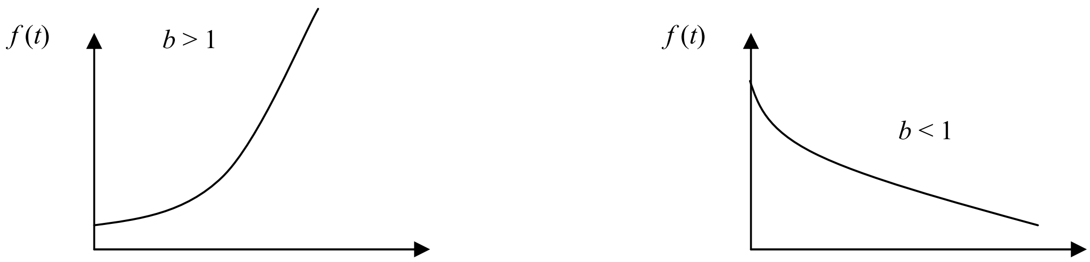


Рис. 7.3. Показательная кривая роста $f(t) = a \cdot b^t$

Рассмотренные выше кривые роста используются для описания процессов без «насыщения», с монотонно (или кусочно-монотонно) возрастающим или убывающим характером тенденции.

Когда процесс характеризуется «насыщением», его следует описывать при помощи кривой, имеющей отличную от нуля асимптоту. Примером такой кривой может служить *модифицированная экспонента*

$$f(t) = c + a \cdot b^t \quad (f(t) = c + e^{a_0 + a_1 \cdot t}),$$

где $y = c$ – горизонтальная асимптота.

Если параметр $a < 0$, то асимптота находится выше кривой, если $a > 0$, то ниже (рис. 7.4). При решении экономических задач чаще всего приходится иметь дело с кривой, у которой $a < 0$, $b < 1$. В этом случае рост происходит с замедлением и стремится к некоторому пределу.

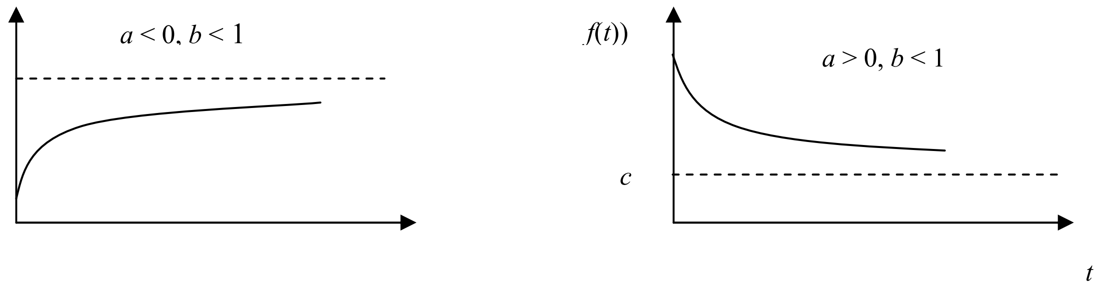


Рис. 7.4. Модифицированная экспонента $f(t)=c+a\cdot b^t$

При решении экономических задач асимптоту кривой роста часто можно определить, исходя из свойств изучаемого процесса. Так, например, коэффициент использования оборудования не может превышать единицы. Иногда значение асимптоты задается экспертным путем. Например, главный инженер предприятия указывает, что производственные мощности не позволяют наращивать объемы производства выше определенного уровня. Этот уровень является тогда оценкой значения асимптоты при прогнозировании производства продукции.

Исследование динамики социальных и экономических процессов выявило довольно сильную распространенность эффекта «насыщения»: выхода на асимптоту при достижении определенных значений показателей. Если воздействие ограничивающего фактора начинает сказываться только после определенного момента (точки перегиба), до которого процесс развивался по некоторому экспоненциальному закону, то для описания такого процесса используют *S*-образные кривые. Наиболее известные из них – кривая Гомперца и логистическая кривая.

Уравнение *кривой Гомперца* имеет вид:

$$f(t)=c\cdot a^{b^t} .$$

Если $\ln a < 0$, кривая имеет *S*-образный вид, при этом асимптота $y = c$ проходит выше кривой. Если $\ln a > 0$, асимптота $y = c$ лежит ниже кривой, а сама кривая меняется монотонно: при $b < 1$ она монотонно убывает, а при $b > 1$ – монотонно возрастает (рис. 7.5).

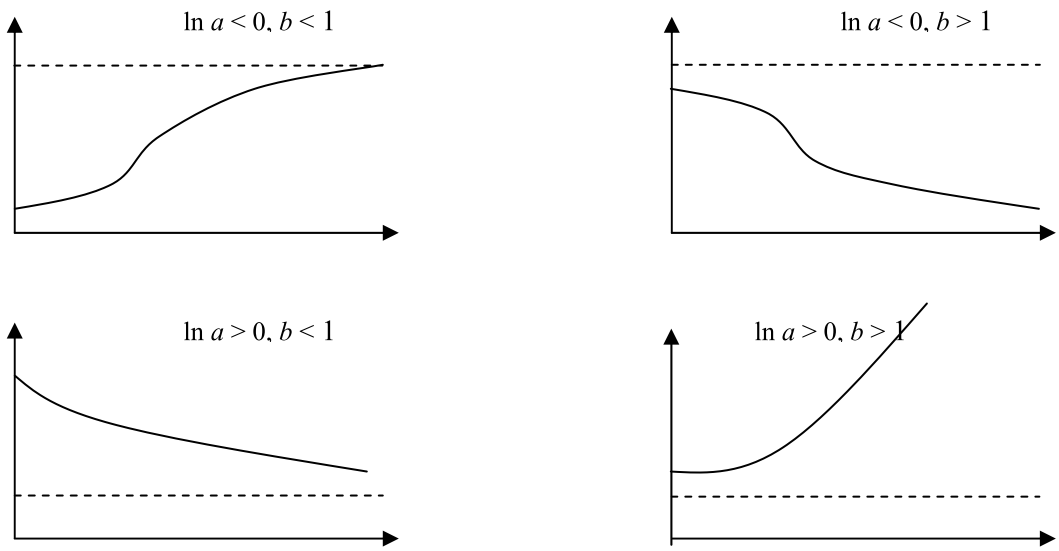


Рис. 7.5. Кривая Гомперца $f(t)=c\cdot a^{b^t}$

К этому же типу кривых роста с «насыщением» относится *логистическая кривая*

$$f(t)=\frac{c}{1+b\cdot e^{-a\cdot t}} .$$

Эта кривая характеризует развитие показателя во времени, когда ускоренный рост в начале периода сменяется замедляющимся темпом роста вплоть до полной остановки, что на графике соответствует отрезку кривой, параллельному оси абсцисс (рис. 7.6).

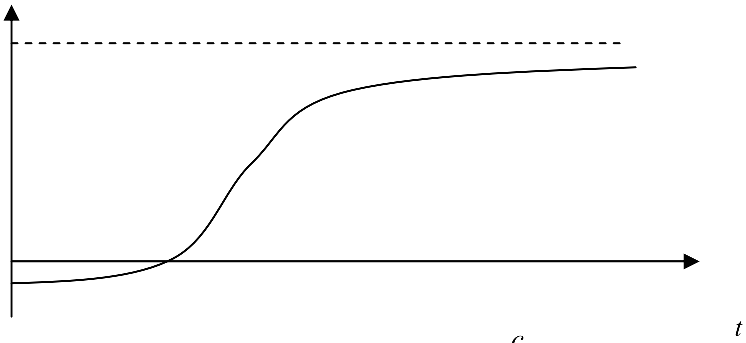


Рис. 7.6. Логистическая кривая $f(t)=\dfrac{c}{1+b\cdot e^{-a\cdot t}}$

С помощью логистической кривой хорошо описывается развитие новой отрасли (нового производства). Сначала технические методы производства еще недостаточно разработаны, издержки производства высоки и спрос на рынке на данный товар еще мал, поэтому производство развивается медленно. В дальнейшем, благодаря усовершенствованию технических методов изготовления, переходу к массовому производству и увеличению емкости рынка для данного товара производство растет быстрее. Затем наступает период насыщения рынка, рост производства все более замедляется и наконец прекращается. Наступает стабилизация производства на определенном уровне.

Для оценки параметров основной тенденции чаще всего используется метод наименьших квадратов (МНК), рассмотренный в темах 2 и 3. При этом значения временного ряда y_t рассматриваются как зависимая переменная, а время t – как объясняющая переменная:

$$y_t = f(t) + \varepsilon_t,$$

где ε_t – «возмущения», удовлетворяющие основным предпосылкам регрессионного анализа, приведенным в п. 3.2.

Напомним, например, что согласно МНК оценки параметров прямой $y_t^* = f(t) = a_0 + a_1 t$ находятся по формулам (2.7) и (2.8) (тема 2), в которых в качестве x_i берется t:

$$a_1^* = \frac{\overline{y_t \cdot t} - \overline{y_t}\cdot\bar{t}}{\overline{t^2} - \bar{t}^2}, \tag{7.7}$$

$$a_0^* = \overline{y_t} - a_1^* \cdot \bar{t}, \tag{7.8}$$

где

$$\overline{y_t} = \frac{1}{n}\sum_{t=1}^{n} y_t; \quad \overline{y_t \cdot t} = \frac{1}{n}\sum_{t=1}^{n} y_t t; \tag{7.9}$$

$$\bar{t} = \frac{1}{n}\sum_{t=1}^{n} t = \frac{n+1}{2}; \quad \overline{t^2} = \frac{1}{n}\sum_{t=1}^{n} t^2 = \frac{(n+1)(2n+1)}{6}. \tag{7.10}$$

В двух последних соотношениях мы воспользовались тем, что значения переменной t = 1, 2, …, n образуют натуральный ряд чисел от 1 до n, и поэтому суммы $\sum_{t=1}^{n} t, \sum_{t=1}^{n} t^2$ можно выразить через число членов ряда n по известным формулам:

$$\sum_{t=1}^{n} t = \frac{n(n+1)}{2}; \quad \sum_{t=1}^{n} t^2 = \frac{n(n+1)(2n+1)}{6}.$$

Пример 7.3. По данным примера 7.1 найти уравнение неслучайной составляющей (тренда) для временного ряда y_t, полагая его линейным. Проверить значимость уравнения тренда на 5%-м уровне.

Решение. По формулам (7.9) и (7.10) рассчитаем средние

$$\overline{y_t}=\frac{1}{8}(213+171+...+361)=296{,}875;$$

$$\overline{y_t\cdot t}=\frac{1}{8}(213\cdot 1+171\cdot 2+...+361\cdot 8)=1470{,}75;$$

$$\bar{t}=\frac{8+1}{2}=4{,}5;\quad \overline{t^2}=\frac{(8+1)(2\cdot 8+1)}{6}=25{,}5\,.$$

Подставляя найденные средние в формулы (7.7) и (7.8), получаем:

$$a_1^*=\frac{1470{,}75-296{,}875\cdot 4{,}5}{25{,}5-4{,}5^2}=25{,}68;\ \ a_0^*=296{,}875-25{,}68\cdot 4{,}5=181{,}32\,,$$

откуда уравнение тренда

$$y_t^*=181{,}32+25{,}68t\,.$$

На рис. 7.7 график трендовой составляющей временного ряда изображен сплошной прямой. Полученное уравнение тренда показывает, что спрос ежегодно увеличивается в среднем на 25,7 усл. ед.

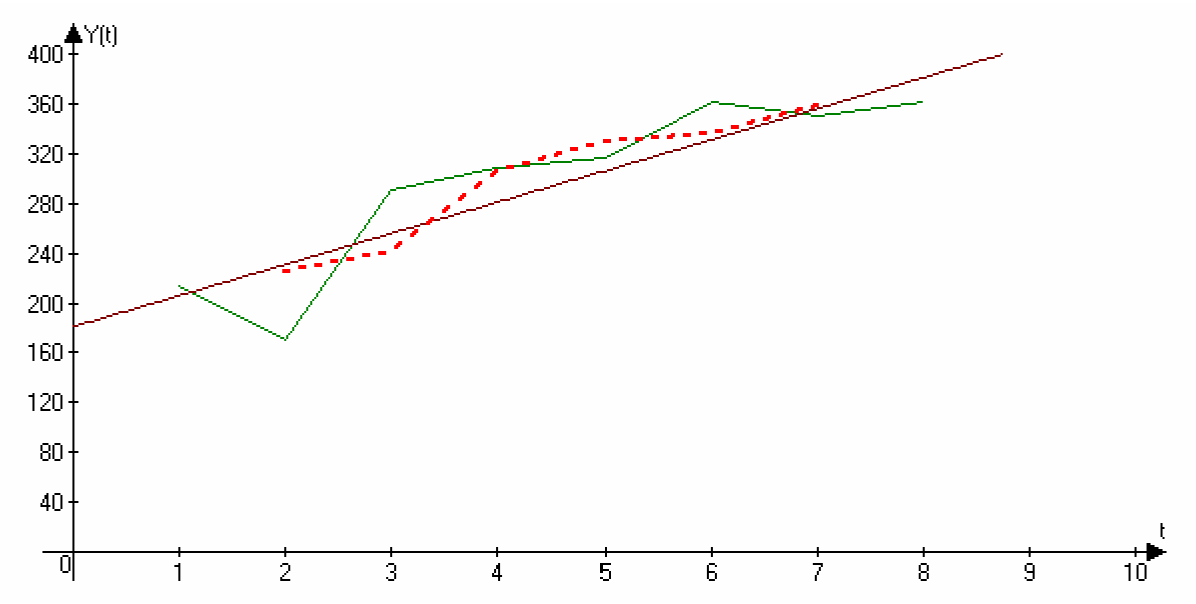


Рис.7.7. Временной ряд (сплошная ломаная), тренд (сплошная прямая) и сглаженный ряд (пунктирная ломаная)

Проверим значимость полученного уравнения тренда по F-критерию на 5%-м уровне значимости. Для этого вычислим вначале сумму квадратов отклонений, обусловленную трендом. В силу соотношения (2.13) уравнение тренда можно записать в виде $y_t^*-\overline{y_t}=a_1^*(t-\bar{t})$. Следовательно,

$$\sum_{t=1}^{n}\left(y_t^*-\overline{y_t}\right)^2=\sum_{t=1}^{n}\left(a_1^*\right)^2(t-\bar{t})^2=\left(a_1^*\right)^2\left(\sum_{t=1}^{n}t^2-\left(\sum_{t=1}^{n}t\right)^2/n\right)=\ 25{,}68^2\left(204-\frac{36^2}{8}\right)=27697{,}42\,.$$

Общая сумма квадратов отклонений равна

$$\sum_{t=1}^{n}\left(y_t-\overline{y_t}\right)^2=\sum_{t=1}^{n}y_t^2-\left(\sum_{t=1}^{n}y_t\right)^2/n=\ 739827-\frac{2375^2}{8}=34748{,}88\,.$$

Найдем по формуле (2.29) коэффициент детерминации: $R^2=\dfrac{27697{,}42}{34748{,}88}=0{,}797$, а по формуле (2.28) – значение F-статистики:

$$F = \frac{0{,}797 \cdot (8-2)}{1-0{,}797^2} = 13{,}11.$$

По табл.3 Приложения находим $F_{1-\alpha}(1, n-2) = F_{0,95}(1,6) = 5{,}99$. Так как $F >$ $F_{1-\alpha}(1, n-2)$ (13,11 > 5,99), то уравнение тренда значимо. ■

Замечание 7.1. При применении метода наименьших квадратов для оценки параметров экспоненциальной, логистической функций или функции Гомперца возникают сложности с решением получаемой системы нормальных уравнений. Поэтому предварительно, до получения соответствующей системы, прибегают к некоторым преобразованиям этих функций (например, логарифмированию и др. (см. п. 4.1)). ◂

7.3.2. АЛГОРИТМИЧЕСКОЕ ВЫРАВНИВАНИЕ ВРЕМЕННОГО РЯДА

При *алгоритмическом выравнивании* (*сглаживании*) временного ряда не предполагается описание динамики неслучайной составляющей с помощью той или иной функции. Процедуры данного метода предоставляют исследователю лишь алгоритм расчета неслучайной составляющей в любой заданный момент времени t. Методы сглаживания временных рядов с помощью скользящих средних относятся к этому подходу.

Скользящие средние позволяют сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии процесса и поэтому служат важным инструментом исследования временных рядов. Метод скользящих средних основан на переходе от начальных значений временного ряда к их средним значениям (простым или взвешенным) на интервале времени, длина которого определена заранее. При этом сам выбранный интервал времени «скользит» вдоль ряда.

Получаемый таким образом ряд скользящих средних ведет себя более гладко, чем исходный ряд, из-за усреднения отклонений ряда. Для усреднения могут быть использованы средняя арифметическая (простая и с некоторыми весами), медиана и др.

Алгоритм сглаживания *по простой скользящей средней* может быть представлен в виде следующей последовательности шагов.

последовательных уровней ряда ($k < n$). При этом надо иметь в виду, что чем шире интервал сглаживания, тем в большей степени поглощаются колебания, и тенденция развития носит более плавный, сглаженный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания.

2. Разбивают весь период наблюдений на отрезки с длиной интервала сглаживания, при этом интервал сглаживания «скользит» по ряду с шагом, равным единице.

3. Рассчитывают средние арифметические из уровней ряда, образующих каждый отрезок.

4. Заменяют фактические значения ряда, стоящие в центре каждого отрезка, на соответствующие средние значения.

При этом удобно брать длину интервала сглаживания κ в виде нечетного числа $k = 2p + 1$, так как в этом случае полученные значения скользящей средней приходятся на средний член интервала.

Пример 7.4. Провести сглаживание временного ряда y_t по данным примера 7.1 методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания $k = 3$ года.

Решение. Скользящие средние находим по формуле:

$$\tilde{y}_t = \frac{1}{k} \sum_{i=t-p}^{t+p} y_i, \tag{7.11}$$

когда $k = 2p + 1$ – нечетное число; при $k = 3$ получаем $p = 1$.

По формуле (7.11) находим:

при $t = 2$

$$\tilde{y}_2 = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}(213 + 171 + 291) = 225;$$

при $t = 3$

$$\tilde{y}_3 = \frac{1}{3}(y_2 + y_3 + y_4) = \frac{1}{3}(171 + 291 + 309) = 241 \text{ и т. д.}$$

В результате получим сглаженный ряд:

t	1	2	3	4	5	6	7	8
$\tilde{y}_t$	–	225,0	241,0	305,7	329,3	336,3	358,0	–

На рис. 7.7 этот ряд изображен графически в виде пунктирной линии. ■

Моделирование сезонных и циклических колебаний

Под *сезонными* (*циклическими*) колебаниями понимаются устойчивые изменения внутри отдельно взятого периода. Часто сезонные колебания связаны со сменой времен года, и поэтому они повторяются ежегодно. Так от времени года существенно зависят потребление топлива, производство сельскохозяйственных продуктов и пр.

Моделирование циклических колебаний в целом осуществляется аналогично моделированию сезонных колебаний, поэтому мы рассмотрим только моделирование последних.

Существует несколько подходов к моделированию сезонных (циклических) колебаний:

- расчет значений сезонной компоненты и построение аддитивной или мультипликативной модели временного ряда;
- применение сезонных фиктивных переменных;
- использование рядов Фурье и др.

Поскольку для целей прогнозирования наиболее важным является применение рядов Фурье, остановимся подробнее на этом методе.

С помощью *ряда Фурье* можно представить *периодические* (сезонные и циклические) колебания $f(t)$, свойственные динамике многих экономических явлений, в виде функции времени t:

$$f(t) = a_0 + \sum_{k=1}^{m} \left(a_k \cdot \cos kt + b_k \cdot \sin kt \right),$$

где a_0, a_k, b_k ($k = 1, 2, \ldots, m$) – параметры ряда Фурье;

$\cos kt$, $\sin kt$ – тригонометрические функции (или *гармоники* ряда Фурье);

k – номер гармоники; m – число гармоник ряда Фурье.

На основе МНК параметры ряда Фурье определяются по следующим формулам:

$$a_0 = \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t; \ a_k = \frac{2}{n}\sum_{t=1}^{n} y_t \cdot \cos kt; \ b_k = \frac{2}{n}\sum_{t=1}^{n} y_t \cdot \sin kt,$$

$$k = 1, 2, \ldots, m.$$

Используя ряд Фурье при изучении сезонности, часто n принимают равным числу месяцев в году, т.е. $n = 12$. Тогда для различных месяцев года t принимает следующие значения (см. табл. 7.4).

Т а б л и ц а 7.4

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь
t	0	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{5\pi}{6}$
Месяц	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
t	π	$\frac{7\pi}{6}$	$\frac{4\pi}{3}$	$\frac{3\pi}{2}$	$\frac{5\pi}{3}$	$\frac{11\pi}{6}$

Полагая число гармоник k, равным 1, 2 и т.д., находят все значения $\cos kt$ и $\sin kt$. Тогда, например, первая гармоника ряда Фурье примет вид:

$$f(t) = a_0 + a_1 \cos t + b_1 \sin t\,.$$

Здесь

$$a_0 = \frac{1}{12}\sum_{t=1}^{12} y_t; \; a_1 = \frac{1}{6}\sum_{t=1}^{12} y_t \cdot \cos t; \; b_1 = \frac{1}{6}\sum_{t=1}^{12} y_t \cdot \sin t \,.$$

Ряд Фурье с двумя гармониками

$$f(t) = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t \,,$$

где

$$a_2 = \frac{1}{6}\sum_{t=1}^{12} y_t \cdot \cos 2t; \; b_2 = \frac{1}{6}\sum_{t=1}^{12} y_t \cdot \sin 2t \; \text{ и т.д.}$$

Пример 7.5. В табл. 7.5 приведены данные (в тыс. усл. ед.) о продаже женской летней обуви по одному району города N.

Таблица 7.5

Месяц	Янв.	Фев.	Март	Апр.	Май	Июнь
Продано обуви	37,0	40,0	44,0	52,0	46,0	70,0
Месяц	Июль	Авг.	Сент.	Окт.	Нояб.	Декаб.
Продано обуви	70,0	48,0	46,0	38,0	36,0	35,0

Представить сезонные колебания с помощью ряда Фурье по одной и двум гармоникам. Изобразить графически исходный временной ряд, а также ряды, сглаженные по одной и двум гармоникам.

Решение. Промежуточные результаты, необходимые для расчета параметров ряда Фурье, приведены в табл. 7.6.

Таблица 7.6

Месяц	t	Продано обуви y_t	$y_t \cdot \cos t$	$y_t \cdot \sin t$	$f_1(t)$	$y_t \cdot \cos 2t$	$y_t \cdot \sin 2t$	$f_2(t)$
1	0	37,0	37,0	0	35,0	37,0	0	37,9
2	$\pi/6$	40,0	34,64	20,0	39,3	20,0	34,64	39,6
3	$\pi/3$	44,0	22,0	38,1	45,5	–22,0	38,1	43,0
4	$\pi/2$	52,0	0	52,0	51,7	–52,0	0	43,8
5	$2\pi/3$	46,0	–23,0	39,84	56,5	–39,84	–34,84	56,2
6	$5\pi/6$	70,0	–60,62	35,0	58,4	35,0	–60,6	61,0
7	π	60,0	–60,0	0	57,0	60,0	0	54,9
8	$7\pi/6$	48,0	–41,57	–24,0	52,7	24,0	41,57	53,0
9	$4\pi/3$	46,0	–23,0	–39,84	46,5	–23,0	39,84	44,0
10	$3\pi/2$	38,0	0	–38,0	39,3	–38,0	0	36,4
11	$5\pi/3$	36,0	18,0	–31,17	35,5	–18,0	–31,18	35,2
12	$11\pi/6$	35,0	30,31	–17,5	33,6	17,5	–30,31	36,2
Σ	–	552,0	–66,24	34,43	551,0	17,5	–7,78	551,2

Параметры первой гармоники ряда Фурье будут следующие:

$$a_0 = \frac{552}{12} = 46; \; a_1 = \frac{-66,24}{6} = -11,04; \; a_2 = \frac{34,43}{6} = 5,74 \,,$$

а уравнение самой гармоники

$$f_1(t) = 46 - 11{,}04\cos t + 5{,}74\sin t\,.$$

Значения сглаженного по первой гармонике временного ряда приведены в 6-м столбце табл. 7.6. Для второй гармоники ряда Фурье дополнительные параметры следующие:

$$a_2 = \frac{17{,}5}{6} = 2{,}9;\; b_2 = \frac{-7{,}78}{6} = -1{,}3\;;$$

уравнение гармоники

$$f_2(t) = 46 - 11{,}04\cos t + 5{,}74\sin t + 2{,}9\cos 2t - 1{,}3\sin 2t\,.$$

Графическое представление временного ряда продаж женской летней обуви, а также сглаженные ряды при помощи первой и второй гармоник ряда Фурье приведены на рис. 7.8.

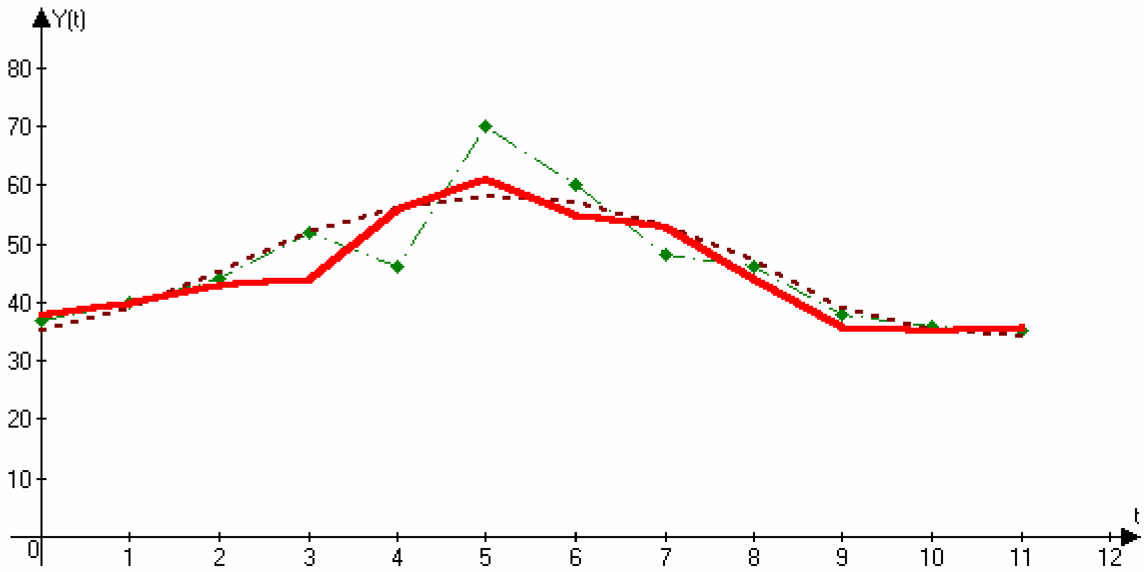


Рис. 7.8. Исходный и сглаженные временные ряды

– исходный временной ряд;

– сглаженный по 1-й гармонике;

– сглаженный по 2-й гармонике)

7.5. ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ

Напомним, что построение модели временного ряда – не самоцель, а лишь средство его анализа и *прогнозирования*. При этом исходят из того предположения, что тенденция развития изучаемого процесса, установленная в прошлом, может быть распространена (*экстраполирована*) на будущий период.

Задача ставится следующим способом: имеется временной ряд y_t (t = 1, 2, …, n) и требуется дать прогноз уровня этого ряда на момент $t + l$.

Один из распространенных подходов к прогнозированию состоит в следующем:

• временной ряд раскладывается на трендовую (долговременную), сезонную (объединяющую в себе собственно сезонную и циклическую компоненты) и случайную компоненты;

• трендовую составляющую «подгоняют» полиномом, а сезонную – рядом Фурье;

• прогноз производится экстраполяцией этих «подогнанных» значений в будущее.

Выше (в п.п. 2.4.4 и 3.3.4) мы рассматривали *точечный* и *интервальный прогноз* значений зависимой переменной Y для значений объясняющих переменных X, расположенных вне пределов обследованного диапазона значений X. Если рассматривать временной ряд как регрессионную модель изучаемого признака по переменной «время», то к нему могут быть применены рассмотренные выше методы анализа.

Замечание 7.2. Напомним, что одна из основных предпосылок регрессионного анализа состоит в том, что отсутствует автокорреляция «возмущений» ε_t ($t = 1, 2, \ldots, n$). Поскольку при работе с временными рядами такое допущение часто оказывается неверным, необходимо проверить отсутствие /наличие автокорреляции, например, с помощью критерия Дарбина-Уотсона (см. п. 5.2.2). ◄

Пример 7.6. По данным примера 7.1 дать с надежностью 0,95 точечный и интервальный прогноз среднего и индивидуального значений спроса на товар на 9-й год, предполагая, что тренд линейный.

Решение. Выясним вначале отсутствие /наличие автокорреляции «возмущений» на уровне значимости $\alpha = 1 - 0{,}95 = 0{,}05$.

Выше, в примере 7.3, получено уравнение регрессии $y_t^* = 181{,}32 + 25{,}68t$. В табл. 7.7 приведен расчет сумм, необходимых для вычисления d-статистики Дарбина-Уотсона.

Т а б л и ц а 7.7

t	y_t	y_t^*	$e_t = y_t - y_t^*$	e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
1	213	207,0	6,0	–	–	36,0
2	171	232,7	–61,7	6,0	4 583,3	3 806,9
3	291	258,4	32,6	–61,7	8 892,5	1 062,8
4	309	284,0	25,0	32,6	57,8	625,0
5	317	309,7	7,3	25,0	313,3	53,3
6	362	335,4	26,6	7,3	372,5	707,6
7	351	361,1	–10,1	26,6	1 346,9	102,0
8	361	386,7	–25,7	–10,1	243,4	660,5
Σ	2 014	–	0	–	15 809,7	7 054,1

Теперь по формуле (5.7) статистика Дарбина-Уотсона

$$d = \frac{15809{,}7}{7054{,}1} = 2{,}24\,.$$

По табл.4 Приложения при уровне значимости $\alpha = 0{,}05$, $m = 1$ (одна объясняющая переменная – время) и объеме выборки $n = 8$ находим критические значения $d_н = 0{,}763$, $d_в = 1{,}332$. Так как $d_в \le d < 4 - d_в$ ($1{,}332 \le 2{,}24 < 2{,}668$), то для рассматриваемого временного ряда принимается гипотеза об отсутствии автокорреляции на уровне значимости 0,05.

Точечный прогноз среднего спроса на товар на момент $t = 9$ найдем с помощью уравнения регрессии

$$y_9^* = 181{,}32 + 25{,}68 \cdot 9 = 412{,}4 \text{ (усл. ед.).}$$

С помощью формулы (2.17) находим оценку дисперсии σ^2 «возмущений»:

$$S^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2} = \frac{7054{,}1}{8-2} = 1175{,}7\,.$$

Вычислим оценки дисперсии и среднеквадратического отклонения для выборочного уравнения регрессии при $t = 9$:

$$S^2_{y_9^*} = S^2\left(\frac{1}{n} + \frac{(t-\bar{t})^2}{\sum\limits_{t=1}^{n}(t-\bar{t})^2}\right) = 1175{,}7\left(\frac{1}{8} + \frac{(9-4{,}5)^2}{42}\right) = 713{,}8\,;$$

$$S_{y_9^*} = \sqrt{713{,}8} = 26{,}72 \text{ (усл. ед.).}$$

Здесь мы использовали данные, полученные в примере 7.3:

$$\bar{t} = 4{,}5;\ \sum_{t=1}^{n}(t-\bar{t})^2 = \sum_{t=1}^{n} t^2 - n\cdot(\bar{t})^2 = 204 - 8\cdot 4{,}5^2 = 42\,.$$

По табл.1 Приложения находим $t_{1-\alpha/2}(n-2) = t_{0,975}(6) = 2{,}447$.

Воспользовавшись формулой (2.31), определим границы доверительного интервала для среднего значения спроса при t = 9:

$$412{,}4 \pm 2{,}447\cdot 26{,}72\,,$$

или

$$(347{,}02;\ 447{,}78).$$

Таким образом, среднему спросу в 9-м году с надежностью 0,95 будет соответствовать доверительный интервал (347,02 усл.ед.; 447,78 усл.ед.).

Аналогично с помощью формулы (2.32) рассчитываем границы интервала, в котором с надежностью 0,95 будет находиться индивидуальное значение спроса при t = 9:

$$412{,}4 \mp 2{,}447\cdot\sqrt{1157{,}7}\cdot\sqrt{1 + \frac{1}{8} + \frac{(9-4{,}5)^2}{42}}\,,$$

или

$$(306{,}85;\ 517{,}95).$$

Следовательно, в 9-м году возможный спрос на товар с вероятностью 0,95 будет находиться в интервале (306,85 усл. ед.; 517,95 усл.ед.). Нетрудно заметить, что он включает в себя доверительный интервал для среднего спроса.■

Контрольные вопросы

1. В чем суть временного ряда?
2. В чем состоит разница между временным рядом и совокупностью реализаций одной случайной величины?
3. Перечислите основные элементы временного ряда и основные этапы их анализа.
4. Выпишите общий вид аддитивной и мультипликативной модели временного ряда.
5. Как проверить гипотезу о наличии тренда временного ряда?
6. Как определяется автокорреляционная функция и для чег она используется?
7. В чем заключается «сглаживание» временного ряда?
8. Какие методы «сглаживания» временных рядов Вы знаете?
9. Что такое функции роста и какие их типы Вы знаете?
10. Какие методы используются для оценки тренда временного ряда?
11. Объясните назначение скользящих средних. Влияние каких компонент временного ряда устраняется с их помощью?
12. Приведите алгоритм расчета простых скользящих средних.
13. Сколько значений временного ряда теряется при использовании скользящей средней с длиной интервала сглаживания $k = 2p + 1$?
14. Объясните назначение рядов Фурье. Влияние каких компонент временного ряда выделяется с их помощью?
15. Как осуществляется прогнозирование на основе временных рядов?

Контрольные задания

В таблице представлены данные об изменении урожайности пшеницы за 16 лет.

Порядковый номер года	1	2	3	4	5	6	7	8
Урожайность, ц/га	15,3	17,2	18,1	17,3	18,9	17,6	20,9	16,9
Порядковый номер года	9	10	11	12	13	14	15	16
Урожайность, ц/га	17,8	18,9	19,2	18,5	21,6	20,1	18,9	19,7

а) представить графически временной ряд и проверить наличие тренда;

б) найти автокорреляционную функцию для лагов $l = 1,2$;

в) найти уравнение тренда временного ряда, полагая, что он линейный, и проверить его значимость на уровне 0,05.

г) провести сглаживание временного ряда, используя пятилетнюю простую скользящую среднюю.

2. В таблице представлены данные, отражающие динамику роста доходов на душу населения за восьмилетний период:

Порядковый номер года	1	2	3	4	5	6	7	8
Доход на душу населения, усл. ед.	1133	1222	1354	1389	1342	1377	1491	1684

Полагая тренд временного ряда линейным и выполнены условия классической регрессионной модели:

а) найти уравнение тренда и оценить его значимость на уровне 0,05;

б) дать точечный и надежностью 0,95 интервальный прогноз среднего и индивидуального значений доходов на десятый год.

Литература

Основная:

1. *Воронович Н.В.*, *Русин Г.Л.* Эконометрика: Методические указания по выполнению контрольных работ. – Новосибирск, НГУЭУ, 2005.
2. *Доугерти К.* Введение в эконометрику. – М.: ИНФРА-М, 1999.
3. *Практикум* по эконометрике / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001.
4. *Эконометрика* / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002.

Дополнительная:

1. *Айвазян С.А.,Мхитарян В.С.* Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.
2. *Афанасьев В.Н., Юзбашев Н.Н.* Анализ временных рядов и прогнозирование. – М.: Финансы и статистика, 2001.
3. *Бородич С.А.* Эконометрика. – Минск: Новое Знание, 2001.
4. *Джонстон Дж.* Эконометрические методы. – М.: Статистика, 1980.
5. *Кремер Н.Ш., Путко Б.А.* Эконометрика. – М.: ЮНИТИ-ДАНА, 2002.
6. *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс. – М.: Дело, 2000.
7. *Семенов А.Т.* Таблицы вероятностных распределений и квантилей: Учебное пособие. – Новосибирск, НГАЭиУ, 1998.
8. *Семенов А.Т.* Теория вероятностей и математическая статистика: Учебно-методический комплекс. – 2-е изд., перераб. и доп. – Новосибирск: НГУЭУ, 2004.

СЛОВАРЬ ТЕРМИНОВ

Автокорреляция – корреляционная зависимость между текущими значениями некоторой переменной и значениями этой же переменной, сдвинутыми на несколько периодов времени (лагов) назад или вперед.

Автокорреляционная функция – зависимость значений коэффициентов автокорреляции от величины лага (порядка коэффициента автокорреляции).

Адекватность модели – соответствие построенной модели реальному моделируемому объекту или процессу.

Верификация модели – проверка качества найденных параметров модели и самой модели в целом.

Взвешенный метод наименьших квадратов (***ВМНК***) – метод наименьших квадратов для модели с гетероскедастичностью, когда ковариационная матрица возмущений диагональна.

Возмущение – случайная составляющая регрессионной модели, характеризующая отклонение результирующей переменной от функции регрессии.

Временной ряд (***динамический ряд, ряд динамики***) – совокупность значений какой-либо переменной за несколько последовательных моментов или периодов времени.

Выборочное уравнение регрессии – оценка теоретического уравнения регрессии, полученная по выборочным данным.

Гетероскедастичность – непостоянство дисперсий возмущений.

Гомоскедастичность – постоянство дисперсии возмущений.

Двухшаговый метод наименьших квадратов (***ДМНК***) – метод оценки параметров сверхидентифицируемой модели, основанный на двукратном применении МНК.

Доверительная полоса – множество, предназначенное для интервального оценивания неизвестной функции регрессии.

Доверительный интервал – интервал приближенных значений неизвестного параметра, построенный по результатам наблюдений.

Зависимость корреляционная – зависимость (при наличии случайных факторов) между переменными, которые не разделяются на независимые и зависимые.

Зависимость регрессионная – зависимость между значениями независимой переменной (или совокупности независимых переменных) и условным средним (математическим ожиданием) зависимой переменной.

Зависимость статистическая (***стохастическая, вероятностная***) – зависимость между значениями одной переменной и законом распределения другой.

Зависимость функциональная – зависимость между переменными, когда значению одной из них соответствует единственное значение другой.

Идентификация модели – оценка параметров модели.

Идентифицируемость модели – возможность получения однозначно определенных параметров модели, заданной системой одновременных уравнений.

Интервальная оценка – статистическая оценка, в которой решениями служат множества точек пространства оцениваемого параметра или параметрические функции.

Интервальное оценивание – метод математической статистики, предназначенный для построения множества приближенных значений неизвестных параметров.

Ковариационная матрица – матрица, элементами которой являются попарные ковариации компонент случайного вектора.

Ковариация – числовая характеристика совместного распределения двух случайных величин X_1 и X_2 с конечными дисперсиями: $\operatorname{cov}(X_1, X_2) = M\{(X_1 - MX_1)(X_2 - MX_2)\}$.

***Косвенный метод наименьших квадратов* (*КМНК*)** – метод оценки параметров для идентифицируемой системы одновременных уравнений, основанный на МНК.

Коррелограмма – график автокорреляционной функции.

Корреляционная матрица – матрица парных коэффициентов корреляции совокупности случайных величин.

***Корреляционное поле* (*поле рассеяния*)** – графическое изображение статистической зависимости между двумя переменными в виде точек на координатной плоскости.

Корреляционный анализ – совокупность методов, применяемых для установления связи между исследуемыми переменными, структуры и степени тесноты этой связи.

Корреляция – зависимость между случайными переменными, не имеющая, вообще говоря, функционального характера.

Коэффициент автокорреляции – коэффициент корреляции между членами одного и того же ряда, взятыми со сдвигом (лагом), который определяет *порядок коэффициента автокорреляции*.

Коэффициент детерминации – мера качества подгонки регрессионной модели к наблюденным значениям результирующей переменной.

Коэффициент корреляции – мера тесноты линейной зависимости между двумя случайными переменными.

Коэффициент регрессии – коэффициент при независимой переменной в уравнении регрессии.

Коэффициент эластичности – показатель эластичности какой-либо величины (функции), определяемый как логарифмическая производная данной величины (функции) по фактору (аргументу).

Кривая* (*линия*) *регрессии – график функции регрессии одной случайной переменной на другую.

Кривая роста – функция, описывающая основную тенденцию временного ряда, в зависимости от времени.

Критерий Дарбина – Уотсона – правило (тест) для проверки гипотезы о наличии автокорреляции.

Критерий серий – тест для проверки гипотезы о наличии тренда временного ряда.

Лаг – величина сдвига друг относительно друга членов одной последовательности или значений одной функции.

Математическая модель – абстракция реального мира, в которой интересующие исследователя отношения между реальными элементами заменены подходящими отношениями между математическими объектами.

***Метод наименьших квадратов* (*МНК*)** – метод статистического оценивания параметров уравнения регрессии, состоящий в том, что оценка неизвестных параметров определяется из условия минимума суммы квадратов отклонений эмпирических значений объясняющей переменной от значений этой переменной, найденных по уравнению регрессии.

Метод скользящего среднего – метод выделения тренда временного ряда, состоящий в том, что некоторая группа значений временного ряда сглаживается многочленом и в качестве значения тренда в средней точке этой группы выбирается значение многочлена.

МНК-оценки – статистические оценки параметров регрессионной модели, полученные методом наименьших квадратов.

Моделирование – исследование каких-либо явлений, процессов или объектов путем построения и изучения их моделей.

Модель – любое представление (мысленное или условное) какого-либо объекта, процесса или явления («оригинала» данной модели), используемое в качестве его «заместителя», «представителя».

Модель аддитивная – модель, представленная как сумма компонент (составляющих).

Модель временного ряда – математическая модель, построенная по данным, характеризующим один и тот же объект за ряд последовательных моментов (периодов) времени.

Модель мультипликативная – модель, представленная как произведение компонент (составляющих).

Модель пространственная – математическая модель, построенная по данным, характеризующим различные объекты в определенный момент (период) времени.

Надежность оценки (***доверительная вероятность***) – вероятность, с которой доверительный интервал «накрывает» неизвестный параметр.

Неидентифицируемость модели – невозможность оценки параметров структурной формы модели через параметры приведенной формы модели.

Остаточная выборочная дисперсия – статистическая оценка остаточной дисперсии.

Остаточная дисперсия – дисперсия случайного возмущения в регрессионной модели.

Предопределенная переменная – экзогенная переменная или эндогенная переменная с лагом, значения которой определены до рассмотрения соотношений.

Приведенная форма модели – система одновременных уравнений, в которых эндогенные переменные выражены только через экзогенные или предопределенные переменные, а также случайные составляющие.

Производственная функция – экономико-математическая модель зависимости объема выпуска продукции от основных факторов производства.

Производственная функция Кобба-Дугласа – экономико-математическая модель зависимости объема выпуска продукции от затрат капитала и труда.

Ранг – номер изучаемого объекта в некоторой совокупности, упорядоченной (или *ранжированной*) по какому либо принципу.

Ранговая корреляция – корреляция между рангами изучаемых объектов.

Ранговый коэффициент корреляции – величина, измеряющая степень тесноты связи между рангами.

Регрессионная модель – уравнение взаимосвязи переменных, содержащее случайное возмущение с нулевым средним (математическим ожиданием).

Регрессионный анализ – совокупность методов исследования регрессионной зависимости между переменными по статистическим данным.

Регрессия – зависимость среднего значения какой-либо случайной переменной от некоторой другой величины или от нескольких величин.

Регрессор – независимая переменная в регрессионной модели.

Сверхидентифицируемость модели – ситуация, когда на основе параметров приведенной формы модели можно получить несколько значений одного параметра структурной формы модели.

Сглаживание (***выравнивание***) – выделение неслучайной составляющей временного ряда.

Сезонная компонента – составляющая временного ряда, отражающая повторяемость экономических процессов в течение определенного (не очень длительного) периода времени.

Система нормальных уравнений – система уравнений для определения коэффициентов регрессии, получаемая методом наименьших квадратов.

Система одновременных уравнений – совокупность взаимосвязанных регрессионных моделей, в которых одни и те же переменные могут одновременно играть роль (в различных уравнениях системы) объясняемых (результирующих) и объясняющих переменных.

Случайная компонента – составляющая модели (в том числе и временного ряда), отражающая влияние не поддающихся учету и регистрации случайных факторов.

Спецификация модели – выбор переменных модели и вида зависимости между ними.

Статистика – функция от результатов наблюдений.

Статистика Дарбина – Уотсона – статистика для проверки наличия автокорреляции с помощью критерия Дарбина – Уотсона.

Статистика критерия – статистика, используемая для определения статистического критерия.

Статистическая оценка – статистика, служащая приближением для неизвестного значения оцениваемого параметра.

Статистические данные – набор наблюдаемых значений одной или нескольких переменных, характеризующих изучаемое явление или рассматриваемый экономический объект.

Статистический критерий (***тест***) – правило принятия решения о справедливости одной из гипотез.

Структурная форма модели – система одновременных уравнений, составляющих исходную модель.

Теорема Гаусса-Маркова – теорема о свойствах МНК-оценок линейной регрессионной модели.

Точечная оценка – статистическая оценка, в которой решениями служат отдельные точки пространства оцениваемого параметра или параметрической функции.

Тренд – общее направления развития модели.

Тренд временного ряда – систематическое изменение временного ряда, описываемое неслучайной функцией от времени.

Уровень временного ряда – отдельное наблюдение временного ряда.

Уровень значимости критерия – вероятность отвергнуть нулевую (проверяемую) гипотезу, когда она верна.

Уравнение регрессии – уравнение, определяющее функцию регрессии.

Функция регрессии – функция, описывающая изменение условного математического ожидания одной случайной переменной при изменении значений другой или нескольких других переменных.

Фиктивная переменная – искусственная переменная, выражающая состояния (градации) качественного фактора.

Циклическая компонента – составляющая временного ряда, отражающая повторяемость экономических процессов в течение длительных периодов времени.

Экзогенная переменная – внешняя по отношению к модели переменная; ее значения определяются вне модели, и поэтому она считается заданной.

Эластичность – характеристика относительного изменения прироста какой-либо величины (функции) при малых относительных изменениях прироста определяющего эту величину фактора (аргумента).

Эндогенная переменная – переменная, значения которой определяются внутри модели.

ПРИЛОЖЕНИЕ

Таблица 1

Квантили $t_p(n)$ порядка p распределения Стьюдента с n степенями свободы

n \ p	0,750	0,900	0,950	0,975	0,990	0,995	0,999
1	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567	318,3088
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	22,3271
3	7649	6377	3534	3,1824	4,5407	5,8409	10,2145
4	7407	5332	1318	2,7764	3,7469	4,6041	7,1732
5	7267	4759	0150	5706	3,3649	4,0321	5,8934
6	7176	4398	1,9432	4469	3,1427	3,7074	5,2076
7	7111	4149	8946	3646	2,9980	4995	4,7853
8	7076	3968	8595	3060	8965	3554	5008
9	7027	3830	8331	2622	8214	2498	2968
10	6998	3722	8125	2281	7638	1693	1437
11	0,6074	1,3634	1,7959	2,2010	2,7181	3,1058	4,0247
12	6955	3562	7823	1788	6810	0545	3,9296
13	6938	3502	7709	1604	6503	0123	8520
14	6924	3450	7613	1448	6245	2,9768	7874
15	6912	3406	7530	1314	6025	9467	7328
16	6901	3368	7459	1199	5835	9208	6862
17	6892	3334	7396	1098	5669	8992	6458
18	6884	3304	7341	1009	5524	8784	6105
19	6876	3277	7291	0930	5395	8609	5794
20	6870	3253	7247	0860	5280	8453	5518
21	0,6864	1,3232	1,7207	2,0796	2,5177	2,8314	3,5272
22	6858	3212	7171	0739	5083	8188	5050
23	6853	3195	7139	0687	4999	8073	4850
24	6848	3178	7109	0639	4922	7969	4668
25	6844	3163	7081	0595	4851	7874	4502
26	6840	3150	7056	0555	4786	7787	4350
27	6837	3137	7033	0518	4727	7707	4210
28	6834	3125	7011	0484	4671	7633	4082
29	6830	3114	6991	0452	4620	7564	3962
30	6828	3104	6973	0423	4573	7500	3852
40	0,6805	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069
50	6794	2987	6759	0086	4033	6778	2614
60	6786	2958	6706	0003	3901	6603	2317
70	6780	2938	6669	1,9944	3808	6479	2108
80	6776	2922	6641	9904	3739	6387	1953
90	6772	2910	6620	9867	3685	6316	1833
100	6770	2901	6602	9840	3642	6259	1737
200	6757	2858	6524	9719	3451	6006	1315
500	6750	2832	6479	9647	3338	5857	1066
∞	6745	2816	6449	9600	3263	5758	0902

Т а б л и ц а 2

Квантили $\chi^2_p(n)$ порядка p χ^2-распределения с n степенями свободы

n \ p	0,001	0,005	0,01	0,025	0,05	0,1	0,2	0,3	0,4
1	2	3	4	5	6	7	8	9	10
1	$0{,}0^52$	$0{,}0^44$	$0{,}0^32$	0,001	0,004	0,016	0,064	0,148	0,275
2	0,002	0,010	0,020	0,051	0,103	0,211	0,446	0,713	1,022
3	0,024	0,072	0,115	0,216	0,352	0,584	1,005	1,424	1,869
4	0,091	0,207	0,297	0,484	0,711	1,064	1,649	2,195	2,753
5	0,210	0,412	0,554	0,831	1,145	1,610	2,343	3,000	3,655
6	0,381	0,676	0,872	1,237	1,635	2,204	3,070	3,828	4,570
7	0,598	0,989	1,239	1,690	2,167	2,833	3,822	4,671	5,493
8	0,857	1,344	1,646	2,180	2,733	3,490	4,594	4,527	6,423
9	1,153	1,735	2,088	2,700	3,325	4,168	5,380	6,393	7,357
10	1,479	2,156	2,558	2,247	3,940	4,865	6,179	7,267	8,295

Продолжение табл. 2

1	2	3	4	5	6	7	8	9	10
11	1,834	2,603	3,053	3,816	4,575	5,578	6,989	8,148	9,237
12	2,214	3,074	3,571	4,404	5,226	6,304	7,807	9,034	10,182

13	2,617	3,565	4,107	5,009	5,892	7,042	8,634	9,926	11,129
14	3,041	4,075	4,660	5,629	6,571	7,790	9,467	10,821	12,079
15	3,483	4,601	5,229	6,262	7,261	8,547	10,307	11,721	13,030
16	3,942	5,142	5,812	6,908	7,962	9,312	11,152	12,624	13,983
17	4,416	5,697	6,408	7,564	8,672	10,085	12,002	13,531	14,937
18	4,905	6,265	7,015	8,231	9,390	10,865	12,857	14,440	15,893
19	5,407	6,844	7,633	8,907	10,117	11,651	13,716	15,352	16,850
20	5,921	7,434	8,260	9,591	10,851	12,443	14,578	16,266	17,809
21	6,447	8,034	8,897	10,283	11,591	13,240	15,445	17,182	18,768
22	6,983	8,643	9,542	10,982	12,338	14,041	16,314	18,101	19,729
23	7,529	9,260	10,196	11,688	13,091	14,848	17,187	19,021	20,690
24	8,085	9,886	10,856	12,401	13,848	15,659	18,062	19,943	21,652
25	8,649	10,520	11,524	13,120	14,611	16,473	18,940	20,867	22,616
26	9,222	11,160	12,198	13,844	15,379	17,292	19,820	21,792	23,579
27	9,803	11,808	12,879	14,573	16,151	18,114	20,703	22,719	24,544
28	10,391	12,461	13,565	15,308	16,928	18,939	21,588	23,647	25,509
29	10,986	13,121	14,256	16,047	17,708	19,768	22,475	24,577	26,475
30	11,588	13,787	14,953	16,791	18,493	20,599	23,364	25,508	27,442
32	12,811	15,134	16,362	18,291	20,072	22,271	25,148	27,373	29,376
34	14,057	16,501	17,789	19,806	21,664	23,952	26,938	29,242	31,313
36	15,324	17,887	19,233	21,336	23,269	25,643	28,735	31,115	33,252
38	16,611	19,289	20,691	22,878	24,884	27,343	30,537	32,992	35,192
40	17,916	20,707	22,164	24,433	26,509	29,051	32,345	34,872	37,134
45	21,251	24,311	25,901	28,366	30,612	33,350	36,884	39,585	41,995
50	24,674	27,991	29,707	32,357	34,764	37,689	41,449	43,313	46,864
60	31,738	35,535	37,485	40,482	43,188	46,459	50,641	53,809	56,620
70	39,036	43,275	45,442	48,758	51,739	55,329	59,898	63,346	66,396
80	46,520	51,172	53,540	57,153	60,391	64,278	69,207	72,915	76,188
90	54,155	59,196	61,754	65,647	69,126	73,291	78,558	82,511	85,993
100	61,918	67,328	70,065	74,222	77,929	82,538	87,945	92,129	95,808

Продолжение табл. 2

n \ p	0,6	0,7	0,8	0,9	0,95	0,975	0,99	0,995	0,999
1	2	3	4	5	6	7	8	9	10
1	0,708	1,074	1,642	2,706	3,841	5,024	6,635	7,879	10,828
2	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	13,816
3	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	16,266
4	4,045	4,878	5,989	7,779	9,488	11,143	13,277	14,860	18,467
5	5,132	6,064	7,289	9,236	11,070	12,832	15,086	16,750	20,515
6	6,211	7,231	8,558	10,645	12,592	14,449	16,812	18,548	22,458
7	7,282	8,383	9,803	12,017	14,067	16,013	18,475	20,278	24,322
8	8,351	9,524	11,030	13,362	15,507	17,535	20,090	21,955	26,125
9	9,414	10,656	12,242	14,684	16,919	19,023	21,666	23,589	27,877
10	10,473	11,781	13,422	15,987	18,307	20,483	23,209	25,188	29,588
11	11,530	12,899	14,631	17,275	19,675	21,920	24,725	26,757	31,264
12	12,584	14,011	15,812	18,549	21,026	23,336	26,217	28,300	32,909
13	13,636	15,119	16,985	19,812	22,362	24,736	27,688	29,819	34,528
14	14,685	16,222	18,151	21,064	23,685	26,119	29,141	31,319	36,123
15	15,733	17,322	19,311	22,307	24,996	27,488	30,578	32,801	37,697
16	16,780	18,418	20,465	23,542	26,296	28,845	32,000	34,267	39,252
17	17,824	19,511	21,615	24,769	27,587	30,191	33,409	35,718	40,790
18	18,868	20,601	22,760	25,989	28,869	31,526	34,805	37,156	42,312
19	19,910	21,689	23,900	27,204	30,144	32,852	36,191	38,582	43,820
20	20,951	22,775	25,038	28,412	31,410	34,170	37,566	39,997	45,315
21	21,991	23,858	26,171	29,615	32,671	35,479	38,932	41,401	46,797
22	23,031	24,939	27,301	30,813	33,924	36,781	40,289	42,796	48,268
23	24,069	26,018	28,429	32,007	35,172	38,076	41,638	44,181	49,728
24	25,106	27,096	29,553	33,196	36,415	39,364	42,980	45,558	51,179
25	26,143	28,172	30,675	34,382	37,652	40,646	44,314	46,928	52,620
26	27,179	29,246	31,795	35,563	38,885	41,923	45,642	48,290	54,052
27	28,214	30,319	32,912	36,741	40,113	43,194	46,963	49,645	55,476

Окончание табл. 2

1	2	3	4	5	6	7	8	9	10
28	29,249	31,391	34,027	37,916	41,337	44,461	48,278	50,993	56,892
29	30,283	32,461	35,139	39,087	42,557	45,722	49,588	52,336	58,301

30	31,316	33,530	36,250	40,256	43,773	46,979	50,892	53,672	59,703
32	33,331	35,665	38,466	42,585	46,194	49,480	53,486	56,328	62,487
34	35,444	37,795	40,676	44,903	48,602	51,966	56,061	58,964	65,247
36	37,505	39,922	42,879	47,212	50,998	54,437	58,619	61,581	67,985
38	39,564	42,045	45,076	49,513	53,384	56,895	61,162	64,181	70,703
40	41,622	44,165	47,269	51,805	55,758	59,342	63,691	66,766	73,402
45	46,761	49,452	52,729	57,505	61,659	65,410	69,957	73,166	80,077
50	51,892	54,723	58,164	63,167	67,505	71,420	76,154	79,490	86,661
60	62,135	65,226	68,972	74,397	79,082	83,298	88,298	91,952	99,607
70	72,358	75,689	79,715	85,527	90,531	95,023	100,42	104,21	112,32
80	82,566	86,120	90,405	96,578	101,88	106,63	112,33	116,32	124,84
90	92,761	96,524	101,05	107,56	113,14	118,14	124,12	128,30	137,21
100	102,95	106,91	111,67	118,50	124,34	129,56	135,81	140,17	149,45

Т а б л и ц а 4

Критические точки $d_н$ и $d_в$ распределения Дарбина – Уотсона

(n – объем выборки, m – число объясняющих переменных модели) *Уровень значимости* $\alpha = 0,05$

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$
1	2	3	4	5	6	7	8	9
6	0,610	1,400						
7	700	356	0,467	1,896				
8	763	332	359	777	0,368	2,287		
9	824	320	629	699	435	128	0,296	2,388
10	879	320	697	641	525	016	0,376	2,414
11	0,927	1,324	0,658	1,604	0,595	1,928	0,444	2,283
12	0,971	331	812	579	658	864	512	177
13	1,010	340	861	562	715	816	574	094
14	045	330	905	551	767	779	632	030
15	077	361	946	543	814	750	685	1,977
16	106	371	982	539	857	728	734	935
17	133	381	1,015	536	897	710	779	900
18	158	391	046	535	933	696	820	872
19	180	401	074	536	967	685	859	848
20	201	411	100	537	998	676	894	828
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812
22	239	429	147	541	053	664	958	797
23	257	437	168	543	078	660	986	785
24	273	446	188	546	101	656	013	775
25	288	454	206	550	123	654	038	767
26	302	461	224	553	143	652	062	759
27	316	469	240	556	162	651	084	753
28	328	476	255	560	181	650	104	747
29	341	483	270	563	198	650	124	743
30	352	489	284	567	214	650	143	739
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735
32	373	502	309	574	244	650	177	732
33	383	508	321	577	258	651	193	730
34	393	514	333	580	271	652	208	728
35	402	519	343	584	283	653	222	726
36	411	525	354	587	295	654	236	724
37	419	530	364	590	307	655	249	723
38	427	535	373	594	318	656	261	722
39	435	540	382	597	328	658	273	722
40	442	544	391	600	338	659	285	721
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720
50	503	585	462	628	421	674	378	721
55	528	601	490	641	452	681	414	724
60	549	616	514	652	480	689	444	727
65	567	629	536	662	503	696	471	731
70	583	641	554	672	525	703	494	735
75	598	650	571	680	543	709	515	739

Продолжение табл. 4

1	2	3	4	5	6	7	8	9
80	611	662	586	688	560	715	534	743
90	635	679	612	703	589	726	566	751

100	654	694	634	715	613	736	592	758
150	720	746	706	760	693	774	679	788
200	758	778	748	789	738	799	728	810

Продолжение табл. 4

Уровень значимости $\alpha = 0,01$

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$
6	0,390	1,142						
7	433	1,036	0,294	1,676				
8	497	1,003	343	489	0,229	2,102		
9	554	0,998	408	389	279	1,873	0,183	2,433
10	604	1,001	466	333	340	1,733	0,230	2,193
11	0,633	1,010	0,319	1,297	0,396	1,640	0,286	2,030
12	697	023	369	274	449	373	339	1,913
13	738	038	616	261	499	326	391	826
14	776	034	660	234	347	490	441	737
15	811	070	700	232	391	464	488	704
16	844	086	737	232	633	446	332	663
17	874	102	772	233	672	432	374	630
18	902	118	803	239	708	422	613	604
19	928	132	833	263	742	413	630	384
20	932	147	863	271	773	411	683	367
21	0,973	1,161	0,890	1,277	0,803	0,803	0,718	1,334
22	0,997	174	914	284	831	831	748	343
23	1,018	187	938	291	838	838	777	334
24	037	199	960	298	882	882	803	328
25	033	211	981	303	906	906	831	323
26	072	222	1,001	312	928	928	833	318
27	089	233	019	319	949	949	878	313
28	104	244	037	323	969	969	900	313
29	119	234	034	332	988	988	921	312
30	133	263	070	339	1,006	1,006	941	311
31	1,147	1,273	1,083	1,343	1,023	1,023	0,960	1,310
32	160	282	100	332	040	040	979	310
33	172	291	114	338	033	033	996	310
34	184	299	128	364	070	070	1,012	311
35	193	307	140	370	083	083	028	312
36	206	313	133	376	098	098	043	313
37	217	323	163	382	112	112	038	314
38	227	330	176	388	124	124	072	313
39	237	337	187	393	137	137	083	317
40	246	344	198	398	148	148	098	318
45	1,288	1,376	1,243	1,423	1,201	1,201	1,136	1,328
50	324	403	283	446	243	243	203	338
55	336	427	320	466	284	284	247	348
60	383	449	330	484	317	317	283	338
65	407	468	377	500	346	346	313	368
70	429	483	400	513	372	372	343	378
75	448	501	422	529	393	393	368	387
80	466	513	441	541	416	516	390	393
90	496	540	474	563	432	587	429	611
100	522	562	503	583	482	604	462	623
150	611	637	598	631	584	663	571	679
200	664	684	633	693	643	704	633	713

Т а б л и ц а 3

Квантили $\boldsymbol{F}_p(n_1, n_2)$ порядка p F-распределения с n_1 и n_2 степенями свободы $p = 0,9$

n_1 / n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39,8	49,5	53,5	55,8	57,2	58,2	58,9	59,4	59,8	60,1	60,7	61,2	61,7	62,0	62,2	62,5	62,7	63,0	63,3

	6	0	9	3	4	0	1	4	6	9	1	2	4	0	6	3	9	6	3
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,70	5,18	5,18	5,17	5,16	5,15	5,14	5,13
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	3,76
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,10
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74	2,72
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	2,29
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18	2,16
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08	2,06
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	1,97
12	18	81	61	48	39	33	28	24	21	19	15	10	06	2,04	2,01	1,99	1,96	1,93	90
13	14	76	56	43	35	28	23	20	16	14	10	05	01	1,98	1,96	93	90	88	85
14	10	73	52	39	31	24	19	15	12	10	05	01	1,96	94	91	89	86	83	80
15	07	70	49	36	27	21	16	12	09	06	02	1,97	92	90	87	85	82	79	76
16	05	67	46	33	24	18	13	09	06	03	1,99	94	89	87	84	81	78	75	72
17	03	64	44	31	22	15	10	06	03	00	96	91	86	84	81	78	75	72	69
18	01	62	42	29	20	13	08	04	00	1,98	93	89	84	81	78	75	72	69	66
19	2,99	61	40	27	18	11	06	02	1,98	1,96	91	86	81	79	76	73	70	67	63
20	2,97	59	38	25	16	09	04	00	1,96	1,94	89	84	79	77	74	71	68	64	61
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,87	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59
22	95	56	35	22	13	06	2,01	97	93	90	86	81	76	73	70	67	64	60	57
23	94	55	34	21	11	05	1,99	95	92	89	84	80	74	72	69	66	62	59	55
24	93	54	33	19	10	04	98	94	91	88	83	78	73	70	67	64	61	57	53
25	92	53	32	18	09	02	97	93	89	87	82	77	72	69	66	63	59	56	52
26	91	52	31	17	08	01	96	92	88	86	81	76	71	68	65	61	58	54	50
27	90	51	30	17	07	00	95	91	87	85	80	75	70	67	64	60	57	53	49
28	89	50	29	16	06	00	94	90	87	84	79	74	69	66	63	59	56	52	48
29	89	50	28	15	06	1,99	93	89	86	83	78	73	68	65	62	58	55	51	47
30	88	49	28	14	05	1,98	93	88	85	82	77	72	67	64	61	57	54	50	45
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	1,38
60	79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35	1,29
120	75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,19
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,55	1,49	1,42	1,38	1,34	1,30	1,24	1,17	1,00

Продолжение табл. 3

$p = 0,95$

n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
n_2																			
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,56
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,91	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	75	89	49	26	11	3,00	2,91	85	80	75	69	62	54	51	47	43	38	34	30
13	67	81	41	18	03	2,92	83	77	71	67	60	53	46	42	38	34	30	25	21
14	60	74	34	11	2,96	85	76	70	65	60	53	46	39	35	31	27	22	18	13
15	54	68	29	06	90	79	71	64	59	54	48	40	33	29	25	20	16	11	07
16	49	63	24	01	85	74	66	59	54	49	42	35	28	24	19	15	11	06	01
17	45	59	20	2,96	81	70	61	55	49	45	38	31	23	19	15	10	06	01	1,96
18	41	55	16	93	77	66	58	51	46	41	34	27	19	15	11	06	02	1,97	92
19	38	52	13	90	74	63	54	48	42	38	31	23	16	11	07	03	1,98	93	88
20	35	49	10	87	71	60	51	45	39	35	28	20	12	08	04	1,99	1,95	90	84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	30	44	05	82	66	55	46	40	34	30	23	15	07	03	1,98	94	89	84	78
23	28	42	03	80	64	53	44	37	32	27	20	13	05	01	96	91	86	81	76
24	26	40	01	78	62	51	42	36	30	25	18	11	03	1,98	94	89	84	79	73
25	24	39	2,99	76	60	49	40	34	28	24	16	09	01	96	92	87	82	77	71
26	23	37	98	74	59	47	39	32	27	22	15	07	1,99	95	90	85	80	75	69
27	21	35	96	73	57	46	37	31	25	20	13	06	97	93	88	84	79	73	67
28	20	34	95	71	56	45	36	29	24	19	12	04	96	91	87	82	77	71	65
29	18	33	93	70	55	43	35	28	22	18	10	03	94	90	85	81	75	70	64
30	17	32	92	69	53	42	33	27	21	16	09	01	93	89	84	79	74	68	62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Продолжение табл. 3

$p = 0{,}99$

n_1 / n_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,48	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,81	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	50	39	30	4,16	4,01	3,86	3,78	70	62	54	45	36
13	9,07	70	74	5,21	4,86	62	44	30	19	10	3,96	3,82	66	59	51	43	34	25	17
14	8,86	51	56	5,04	69	46	28	14	03	3,94	80	66	51	43	35	27	18	09	00
15	68	36	42	4,89	56	32	14	00	3,89	80	67	52	37	29	21	13	05	2,96	2,87
16	53	23	29	77	44	20	03	3,89	78	69	55	41	26	18	10	02	2,93	84	75
17	40	11	18	67	34	10	3,93	79	68	59	46	31	16	08	00	2,92	83	75	65
18	29	01	09	58	25	01	84	71	60	51	37	23	08	00	2,92	84	75	66	57
19	18	5,93	01	50	17	3,94	77	63	52	43	30	15	00	2,92	84	76	67	58	49
20	10	5,85	4,94	43	10	3,87	70	56	46	37	23	09	2,94	2,86	78	69	61	52	42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2, 64	2,55	2,46	2,36
22	7,93	72	82	31	3,99	76	59	45	35	26	12	2,98	83	75	67	58	50	40	31
23	88	66	76	26	94	71	54	41	30	21	07	93	78	70	62	54	45	35	26
24	82	61	72	22	90	67	50	36	26	17	03	89	74	66	58	49	40	31	21
25	77	57	68	18	85	63	46	32	22	13	2,99	85	70	62	54	45	36	27	17
26	72	53	64	14	82	59	42	29	18	09	96	81	66	58	50	42	33	23	13
27	68	49	60	11	78	56	39	26	15	06	93	78	63	55	47	38	29	20	10
28	64	45	57	07	75	53	36	23	12	03	90	75	60	52	44	35	26	17	06
29	60	42	54	04	73	50	33	20	09	00	87	73	57	49	41	33	23	14	03
30	56	39	51	02	70	47	30	17	07	2,98	84	70	55	47	39	30	21	11	01
40	7,31	5,18	4,31	3,83	3,51	3.29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

СОДЕРЖАНИЕ

Учебное издание

Семенов А.Т., Воронович Н.В.

ЭКОНОМЕТРИКА

Учебно-методический комплекс
для заочной и дистанционной форм обучения

Оператор компьютерной верстки *Л.В. Иванова*

Подписано в печать 00.00.2005 г. Формат 60x84$^1/_8$. Тираж 450 экз.
Гарнитура Times New Roman. Усл. печ. л. 13,5.

Новосибирский государственный университет экономики и управления
6300099, г. Новосибирск, ул. Каменская, 56