МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

по организации лабораторных работ
по дисциплине «Моделирование и оптимизация процессов и систем сервиса»
для студентов направления подготовки
43.03.01 Сервис
Направленность (профиль) «Сервисная экономика»

Ставрополь, 2026 г.

Методические указания по дисциплине «Моделирование и оптимизация процессов и систем сервиса» содержат задания для студентов, необходимые для проведения лабораторных работ.

Проработка практических заданий позволит студентам приобрести необходимые знания в области математических методов и моделирования в управлении и систематизировать знания, полученные на лекциях.

СОДЕРЖАНИЕ

Введение

Курс «Моделирование и оптимизация процессов и систем сервиса» содержит теоретические и практические вопросы применения математических методов и моделей в экономике.

Целью изучения дисциплины «Моделирование и оптимизация процессов и систем сервиса» является получение базовых знаний и основных навыков: по построению экономико-математических моделей; по использованию инструментов компьютерного анализа и моделирования; по созданию расчетных моделей.

Основными целями лабораторных работ является: ознакомление с основами экономико-математического моделирования; определение основных методик целочисленного анализа методов оптимизации, в том числе специальных задач линейного программирования и теории игр.

Практические работы выполняются на основе знаний, полученных при изучении курса на лекционных, индивидуальных и самостоятельных занятиях, а также при изучении смежных дисциплин.

ЛАБОРАТОРНАЯ РАБОТА 1

Предмет и метод моделирования и

оптимизации процессов и систем сервиса

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты.

Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: заключается в том, что эконометрика является наукой, возникшей в результате слияния экономической теории, математики и статистики, причем слияние это принесло синергетический эффект.

Теоретическая часть

Существуют следующие определения эконометрики:

«Эконометрика занимается определением наблюдаемых в экономической жизни конкретных количественных закономерностей, применяя для этой цели статистические методы» (О. Ланге)

«Эконометрика - любое приложение математики или статистических методов к изучению экономических явлений» (Э. Маленво).

Эконометрика - это статистико-математический анализ экономических отношений (Д. Лайтхилл).

Наиболее полное определение дал Р. Фриш в 1933 г.:

«Эконометрика - это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек: статистика, экономическая теория и математика - необходимое, но недостаточное условие для понимания соотношений в современной экономической жизни. Это - единство всех трех составляющих. И это единство образует эконометрику».

Обобщая приведенные выше определения, можно сказать, что эконометрика является наукой, возникшей в результате слияния экономической теории, математики и статистики, причем слияние это принесло синергетический эффект (рис. 1.1).

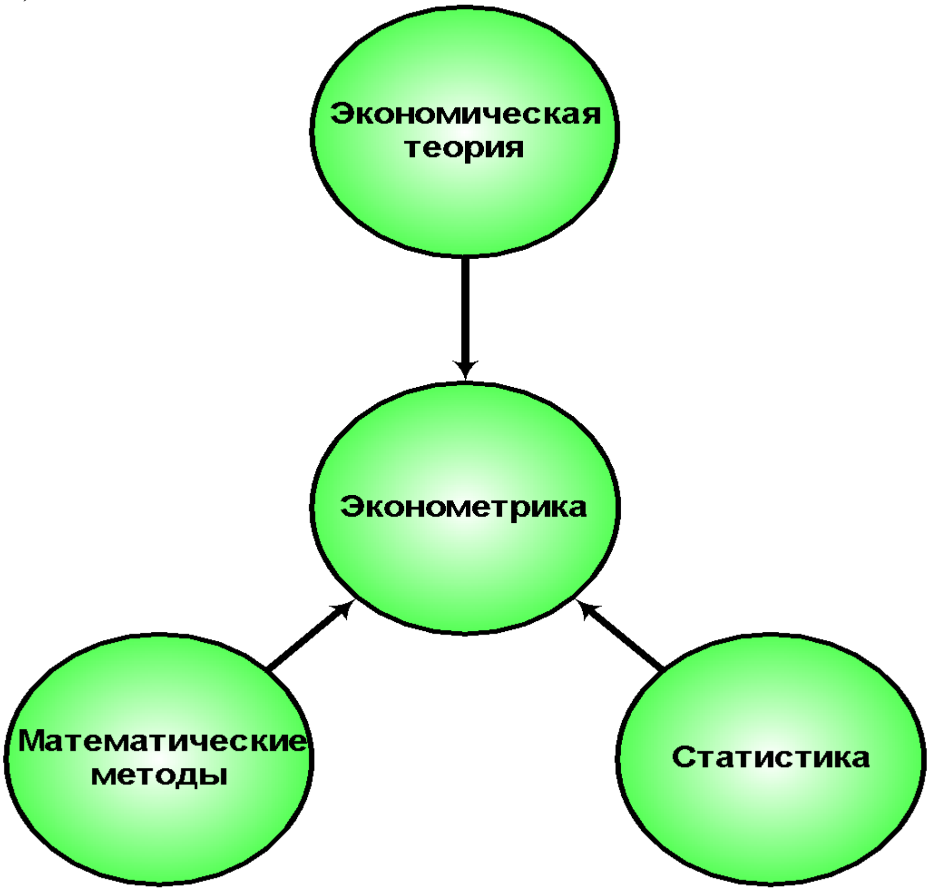


Рис. 1.1 Эконометрика как наука

Исходя из приведенных определений, *предметом изучения* эконометрики является количественное выражение взаимосвязей в экономике.

Первые попытки проведения систематических исследований количественных соотношений в экономике относятся к 17 веку. Одними из первых исследователей, действовавших в этом направлении, были В. Пети и Г. Кинг (школа политических арифметиков). В то время основное внимание уделялось проблеме количественного анализа национального дохода.

Позднее, в 19 веке, в экономической теории возникло маржиналистское направление, активно использующее количественные соотношения и зависимости для объяснения экономических явлений и процессов. Это, а также параллельно проходившее бурное развитие методов статистического и математического анализа обусловило выход эконометрики на качественно иной уровень и окончательное ее формирование как науки.

Годом выделения эконометрики как отдельной науки принято считать 1930 год, когда И. Фишер, Р. Фриш, Я. Тинберген, Й. Шумпетер и др. создали на заседании Американской ассоциации развития науки *эконометрическое общество*.

Методология эконометрики также представляет собой синтез математической, статистической и экономической методологии. Она базируется на общем фундаменте диалектического метода с использованием таких общенаучных методов, как методы индукции и дедукции, научной абстракции, анализа и синтеза, научного эксперимента.

Познание в эконометрике осуществляется в форме *эконометрического исследования*. Его основные этапы представлены на рисунке 1.2.

Рассмотрим последовательно выделенные этапы.

1. Эконометрическое исследование начинается с констатации определенного факта социально-экономической жизни. Это может быть какая-либо *проблема*, например, высокий уровень инфляции, макроэкономическая нестабильность, угроза банкротства, или логически определенная *качественная зависимость*: «объем спроса обратно пропорционален цене», или же продиктованная практическими потребностями *необходимость решения какой-либо задачи*: прогнозирования курсов валют, ценных бумаг и других экономических показателей.

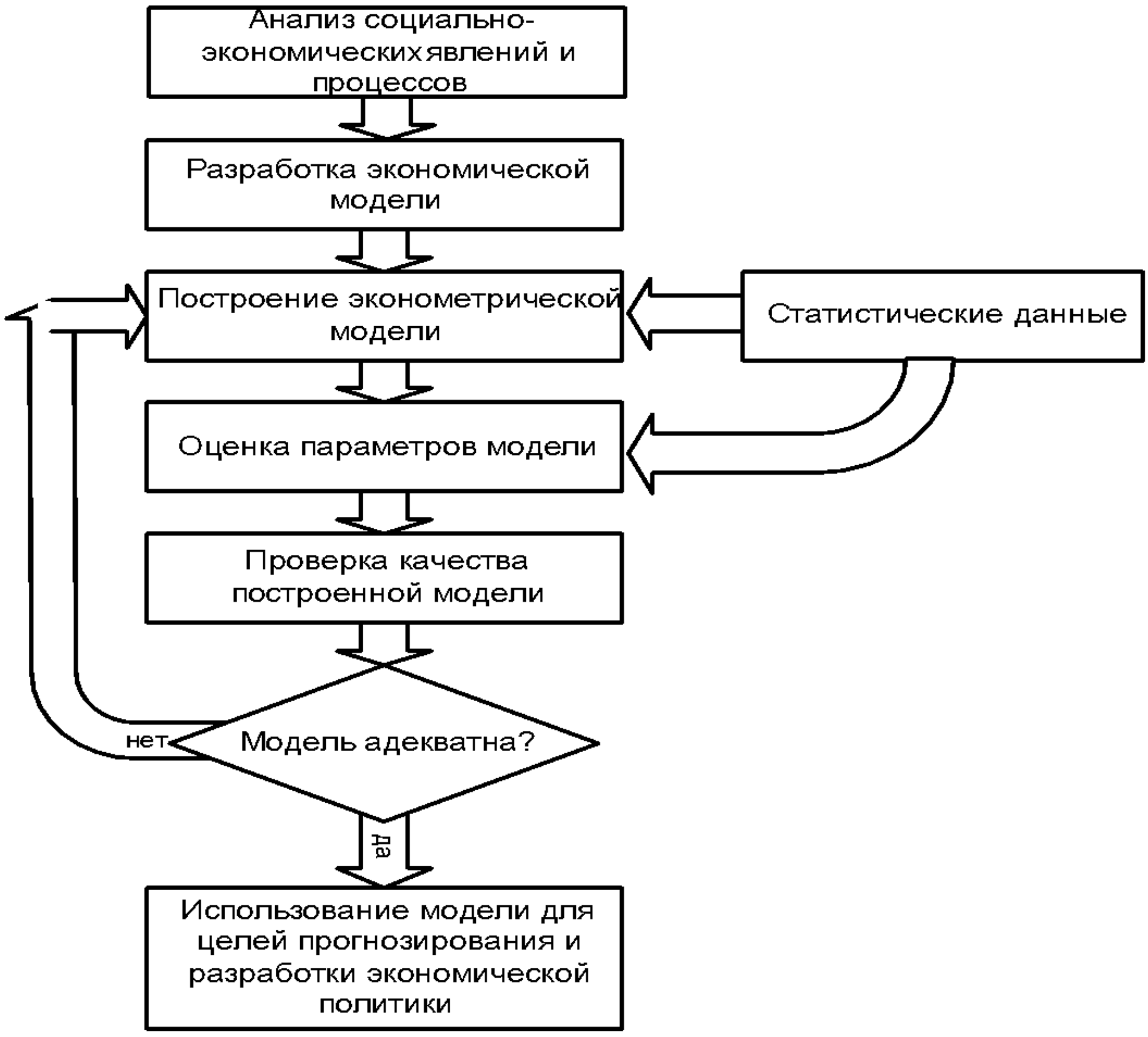


Рис. 1.2 Основные этапы эконометрического исследования

2. На втором этапе на основе метода научной абстракции формируется *описательная экономическая модель*, базирующаяся на выбранных или разработанных исследователем экономических теориях и концепциях (неоклассической, институциональной, кейнсианской, монетаристской и т.п.). Формирование экономической модели не зависит от имеющихся в распоряжении исследователя эмпирических данных.

3. В центре любого эконометрического исследования находится *эконометрическая модель*, разрабатываемая на третьем этапе. Различают два основных подхода к формированию эконометрической модели: «сверху вниз» и «снизу вверх».

Подход *«сверху вниз»* состоит в том, что изначально в модель включается максимально возможное число переменных-факторов. Затем происходит оценка значимости переменных, и те переменные, которые не оказывают существенного влияния на изучаемое явление, исключаются.

При применении подхода *«снизу вверх»*, напротив, изначально выбирается максимально простая модель, содержащая только один основной фактор. Затем, если эта модель недостаточно точна, в нее вводят новые переменные. Процесс усложнения модели осуществляется до тех пор, пока она не будет иметь удовлетворительную точность.

Необходимо отметить, что все переменные можно разделить на *полезные, лишние и вредные*. Полезные – это переменные, введение которых в модель значительно улучшало ее качество. Лишними называются переменные, не оказывающие существенного влияния ни на качество модели, ни на ее параметры, а вредные переменные, в случае их добавления в модель, изменяют значения параметров в ней без существенного изменения качества.

Таким образом, подход «сверху вниз» заключается в обнаружении и исключении из модели лишних и вредных переменных, а подход «снизу вверх» – в поиске и добавлении в модель полезных переменных.

Сравнительный анализ достоинств и недостатков этих подходов приведены на рисунке 1.3

На практике из-за своей сравнительной простоты чаще используется подход «снизу вверх», в то время как серьезное фундаментальное исследование должно основываться на подходе «сверху вниз», поскольку только с его помощью возможен всесторонний анализ явления и установление всех существующих связей.

	«сверху вниз»	«снизу вверх»
достоинства	возможность открытия новых, неожиданных тенденций и закономерностей	простота реализации
недостатки	сложность построения модели, поиска эмпирических данных и расчета параметров	ограниченнность результатов

Рис. 1.3 Достоинства и недостатки основных подходов к построению эконометрической модели

Задания

1. Опишите процесс проведения эконометрического исследования (в соответствии с алгоритмом) следующих проблем:

Микроэкономические проблемы:

- анализ производства в долгосрочном периоде
- рыночное равновесие и его сезонные особенности
- паутинообразная модель восстановления равновесия
- формирование индивидуального спроса

Макроэкономические проблемы

- функция совокупного спроса
- функция совокупных расходов
- модель потребительских ожиданий
- модель государственного регулирования экономического роста

Проблемы рынка ценных бумаг

- динамика курсовой стоимости акций и капитализации компаний
- формирование цен на производные финансовые инструменты
- формирование портфеля ценных бумаг

Проблемы предприятия

- оценка доли фирмы на рынке и объема продаж
- анализ функционирования отдельных подсистем предприятия

2. Объясните на конкретном примере, чем отличаются подходы «сверху вниз» и «снизу вверх».

Литература: [2, 4].

ЛАБОРАТОРНАЯ РАБОТА 2

Предмет и метод моделирования и оптимизации

процессов и систем сервиса

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты.

Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: заключается в том, что эконометрика является наукой, возникшей в результате слияния экономической теории, математики и статистики, причем слияние это принесло синергетический эффект.

Теоретическая часть

Основная проблема, связанная с построением эконометрических моделей, заключается в том, что взаимосвязи между большинством экономических факторов являются не детерминированными, а стохастическими, то есть носящими вероятностный характер. В связи с этим ни одна эконометрическая модель не в состоянии полностью описать экономическую действительность, и неизбежно содержит случайную (необъясненную) компоненту. Эта необъясненная моделью составляющая – ошибка регрессии – говорит о том, что модель не может со стопроцентной точностью спрогнозировать значение результирующей переменной под влиянием объясняющих. Случайная ошибка, как правило, обозначается «ε».

В зависимости от цели исследования и специфики экономической модели выбирается нужный *класс эконометрической модели*. Различают следующие классы:

1) *факторные (регрессионные) статические модели*

$$y = f(x_1, x_2, \dots, x_m) + \varepsilon y = f(x_1, x_2, \dots, x_m) + \varepsilon \quad (1.1)$$

2) *динамические модели*

a. факторные динамические модели (модели с лаговыми переменными):

- модели с лаговыми независимыми переменными:

$$y = f\left(x_1^t, x_2^t, \dots, x_m^t, x_1^{t-1}, x_2^{t-1}, \dots, x_m^{t-1}, \dots, x_1^{t-k}, x_2^{t-k}, \dots, x_m^{t-k}\right) + \varepsilon_t$$

$$y = f\left(x_1^t, x_2^t, \dots, x_m^t, x_1^{t-1}, x_2^{t-1}, \dots, x_m^{t-1}, \dots, x_1^{t-k}, x_2^{t-k}, \dots, x_m^{t-k}\right) + \varepsilon_t \quad (1.2)$$

- авторегрессионные модели:

$$y = f\left(x_1^t, x_2^t, \dots, x_m^t, x_1^{t-1}, \dots, x_1^{t-k}, x_2^{t-k}, \dots, x_m^{t-k}, y^{t-1}, y^{t-2}, \dots, y^{t-k}\right) + \varepsilon_t$$

$$y = f\left(x_1^t, x_2^t, \dots, x_m^t, x_1^{t-1}, \dots, x_1^{t-k}, x_2^{t-k}, \dots, x_m^{t-k}, y^{t-1}, y^{t-2}, \dots, y^{t-k}\right) + \varepsilon_t \quad (1.3)$$

b. модели тренда

$$y = T(t) + \varepsilon_t y = T(t) + \varepsilon_t \qquad (1.4)$$

c. циклические модели (модели сезонности)

$$y = S(t) + \varepsilon_t \, y = S(t) + \varepsilon_t \qquad (1.5)$$

d. модели тренда и сезонности (нестационарные циклические модели):

- аддитивные:

$$y = T(t) + S(t) + \varepsilon_t y = T(t) + S(t) + \varepsilon_t \qquad (1.6)$$

- мультипликативные:

$$y = T(t) \cdot S(t) + \varepsilon_t \, y = T(t) \cdot S(t) + \varepsilon_t \qquad (1.7)$$

e. комбинированные динамические факторные модели

$$y = f(t, x_1^t, x_2^t, \dots, x_m^t) + \varepsilon_t y = f(t, x_1^t, x_2^t, \dots, x_m^t) + \varepsilon_t \qquad (1.8)$$

3) модель системы одновременных уравнений

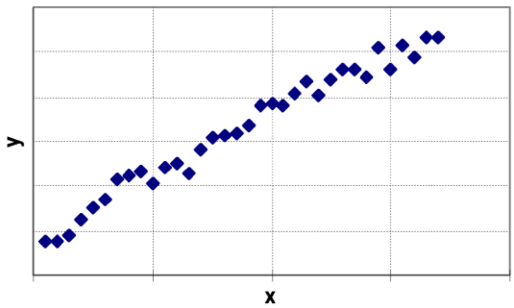
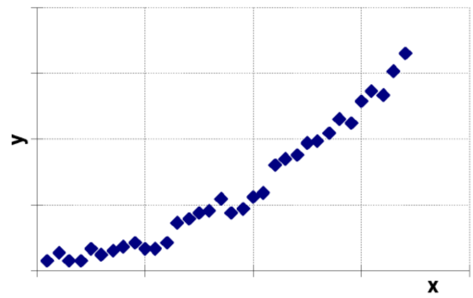
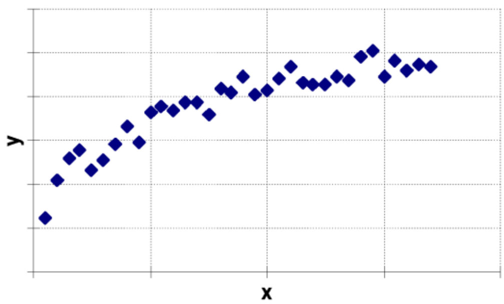
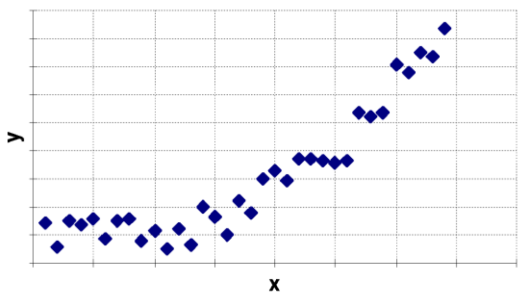
$$\begin{cases} y_1 = f(x_1, x_2, \dots, x_m) + \varepsilon \\ y_2 = f(x_1, x_2, \dots, x_m) + \varepsilon \\ \cdots \\ y_m = f(x_1, x_2, \dots, x_m) + \varepsilon \end{cases} \quad \begin{cases} y_1 = f(x_1, x_2, \dots, x_m) + \varepsilon \\ y_2 = f(x_1, x_2, \dots, x_m) + \varepsilon \\ \cdots \\ y_m = f(x_1, x_2, \dots, x_m) + \varepsilon \end{cases} \qquad (1.9)$$

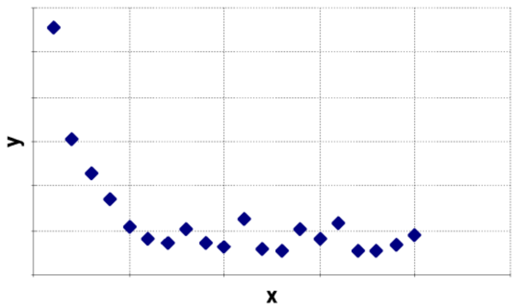
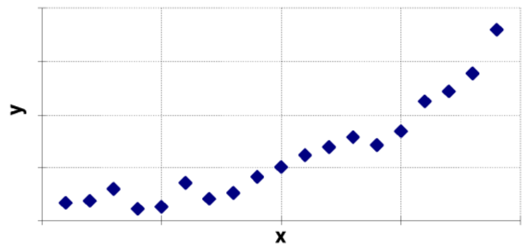
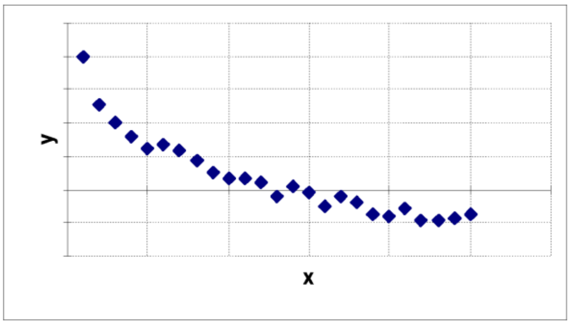
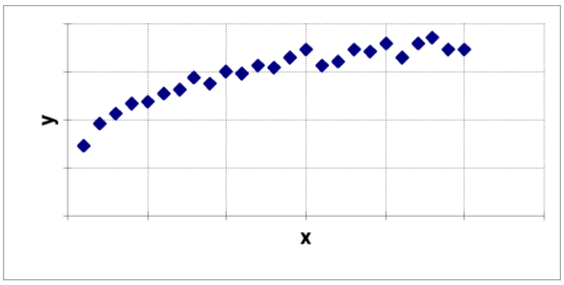
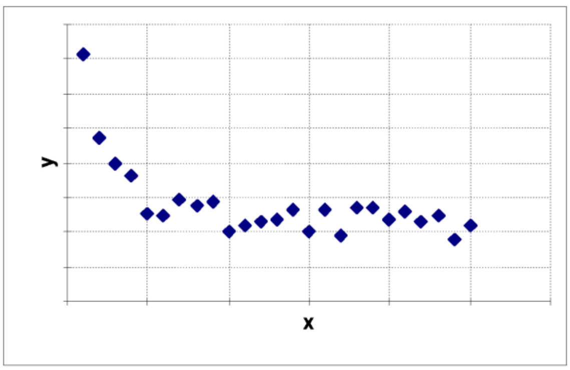
В зависимости от количества изучаемых факторов – объясняющих переменных – (x_i, $i=1..m$) различают парную ($m=1$) и множественную ($m>1$) эконометрическую модель.

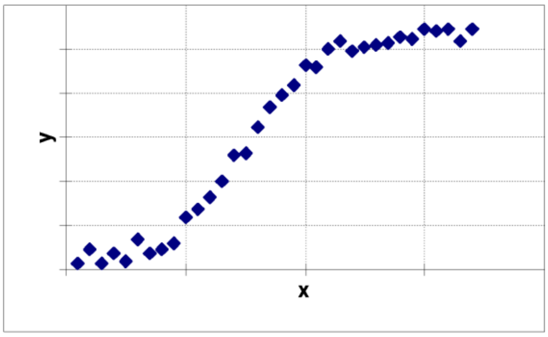
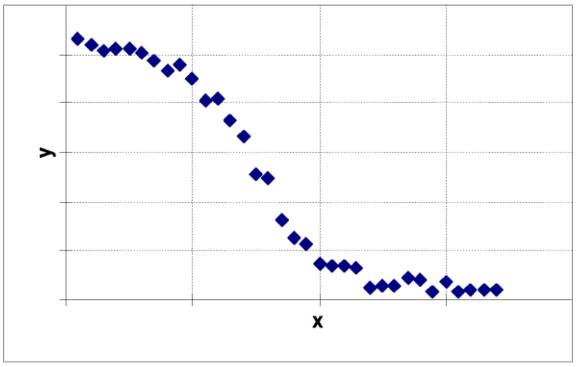
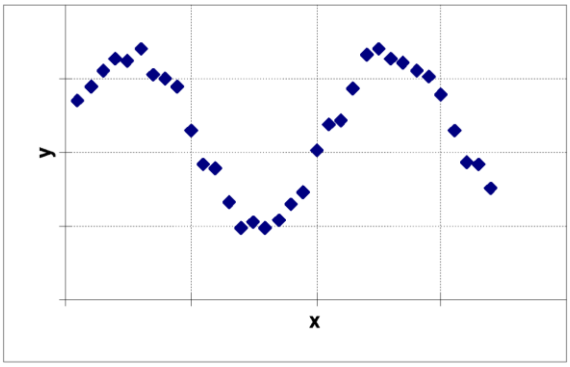
Затем выбирается *функциональная форма эконометрической модели* на основе анализа существующих статистических данных. Выбор формы модели зависит от характера эмпирической зависимости, устанавливаемой либо по корреляционному полю, либо с использованием специальных методов, изучаемых в теории статистики, либо на основании теоретических расчетов.

Для моделей, содержащих только одну объясняющую переменную, основные виды моделей выглядят следующим образом (табл. 1.1)

Таблица 1.1 – Основные функциональные формы эконометрических моделей

№п /п	Вид модели	Аналитическое выражение	Соответствующее корреляционное поле
1.	Линейная	$y = \alpha_0 + \alpha_1 x + \varepsilon$ $y = \alpha_0 + \alpha_1 x + \varepsilon$, где α_0 – начальный уровень *y*, α_1 – скорость изменения *y* при росте *x*	
2.	Параболическая	$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \varepsilon$ где α_0 – начальный уровень *y*, α_1 – скорость изменения *y* при росте *x*, α_2 – величина ускорения изменения *y* при росте *x*	
3.	Степенная	$y = \alpha_0 + \alpha_1 x^{\alpha_2} + \varepsilon$ $y = \alpha_0 + \alpha_1 x^{\alpha_2} + \varepsilon$ где α_0 – начальный уровень *y*, α_1 *и* α_2 – параметры, характеризующие скорость изменения *y* при росте *x*	1. $0 < \alpha_2 < 1$  2. $\alpha_2 > 1$ 

№п/п	Вид модели	Аналитическое выражение	Соответствующее корреляционное поле
4.	Показательная	$y = \alpha_0 + \alpha_1 \cdot \alpha_2^x + \varepsilon$ где $(\alpha_0 + \alpha_1)$ – начальный уровень *y*, α_1 *и* α_2 – параметры, характеризующие скорость изменения *y* при росте *x*	1. $0 < \alpha_2 < 1$  2. $\alpha_2 > 1$ 
5.	Логарифмическая	$y = \alpha_0 + \alpha_1 \cdot \log_{\alpha_2} x + \varepsilon$, где α_0 – уровень *y при x=1*, α_1 *и* α_2 – параметры, характеризующие скорость изменения *y* при росте *x*	1. $0 < \alpha_2 < 1$  2. $\alpha_2 > 1$ 
6.	Гиперболическая	$y = \alpha_0 + \dfrac{\alpha_1}{\alpha_2 + x} + \varepsilon$ где α_0 – уровень *y при x* □ ∞, α_1 –скорость уменьшения *y* при росте *x* α_2 – предельное значение *x*, начиная с которого начинает убывать *y*	

№п/п	Вид модели	Аналитическое выражение	Соответствующее корреляционное поле
7.	Логистическая	$y = \frac{\alpha_0}{1 + e^{\alpha_1 + \alpha_2 x}} + \varepsilon$, где α_0 – предельное значение ряда , α_1 – параметр, характеризующий начальное значение фактора *x*, α_2 – скорость изменения *y*	1. $\alpha_2 < 0$; $\alpha_0 > 0$  1. $\alpha_2 > 0$; $\alpha_0 > 0$ 
8.	Тригонометрическая	$y = \alpha_0 + \alpha_1 \cdot \sin(\alpha_2 x) + \varepsilon$, где α_0 – значение ряда, вокруг которого осуществляется колебательный процесс α_1 – амплитуда колебания *y*, α_2 – скорость изменения *y*	
9.	Комбинированные		

Как видно из таблицы, для описания одних и тех же эмпирических данных вполне могут подойти различные формы моделей. В этом случае выбор полностью зависит от воли исследователя. При прочих равных условиях предпочтение отдается более простым моделям, содержащим наименьшее число параметров (в частности, линейной модели).

4. На следующем этапе происходит *определение параметров эконометрической модели*. Для этого существует целый ряд методов (метод выбранных точек, метод проб, метод наименьших модулей и т.п.), однако наибольшее распространение получили метод наименьших квадратов (МНК) и метод максимального правдоподобия (ММП).

5. *Оценка качества модели.* После определения параметров модели необходимо оценить, насколько хорошо она согласуется с эмпирическими данными, а также насколько существенно значение каждого параметра. При проверке адекватности всей модели используется коэффициент детерминации

(R^2), а для проверки значимости коэффициентов – метод статистической проверки гипотез относительно равенства каждого параметра 0 (или 1). Если гипотеза отвергается, то параметр признается значимым.

В том случае, если признается неудовлетворительное качество модели, производится возврат на третий этап алгоритма, и модель модифицируется: лишние параметры и переменные исключаются и (или) добавляются новые объясняющие переменные.

6. После того, как построенная модель признана адекватной, с ее помощью можно анализировать взаимосвязи между экономическими явлениями и процессами, осуществлять прогнозирование, а также проводить дальнейшие экономические исследования в этой предметной области.

Задание 1

1. Определите, какие модели целесообразно использовать для массивов данных, для которых, параллельно с увеличением факторного признака, характерны:

1) линейный рост результативного признака
2) ускоряющийся рост результативного признака
3) замедляющийся неограниченный рост результативного признака
4) замедляющийся ограниченный рост результативного признака
5) линейное уменьшение результативного признака
6) ускоряющееся уменьшение результативного признака
7) замедляющееся неограниченное уменьшение результативного признака
8) замедляющееся ограниченное уменьшение результативного признака
9) чередование периодов уменьшения и роста результативного признака

По результатам анализа заполните таблицу:

№$_{п/п}$	Характер эмпирических данных	Вид теоретической модели	значения параметров модели
1.	Линейный рост	линейная модель	$a_1 > 0$
		…	…
		…	…
2.	Ускоряющийся рост	…	…
3.	…	…	…

2. Какая из изученных Вами моделей может адекватно описать следующие социально-экономические явления и процессы:

● динамика оборотных фондов быстрорастущего предприятия

- изменение уровня цен при постоянном темпе инфляции
- неоклассическую производственную функцию Кобба-Дугласа
- взаимосвязь между инфляцией и безработицей в краткосрочном периоде
- зависимость потребительских расходов от уровня национального дохода
- изменение национального дохода при увеличении налоговой ставки (см. теорию мультипликатора)
-

Литература: [1, 4, 5].

ЛАБОРАТОРНАЯ РАБОТА 3

Базовые понятия эконометрики

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты.

Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: заключается в том, что большинство величин, с которыми имеет дело исследователь в эконометрике, являются не жестко детерминированными, а вероятностными величинами. Мы можем лишь оценить их значения, предположить с определенной вероятностью границы, в которых они могут находиться на самом деле. Это связано с объективно существующими ошибками наблюдения.

Теоретическая часть

Большинство величин, с которыми имеет дело исследователь в эконометрике, являются не жестко детерминированными, а вероятностными величинами. Мы можем лишь оценить их значения, предположить с определенной вероятностью границы, в которых они могут находиться на самом деле. Это связано с объективно существующими ошибками наблюдения.

Для описания случайных величин используют две основные *характеристики*:

1. ожидаемое значение

2. меру разброса значений (как правило, дисперсию)

Случайная величина записывается в виде: $x = \mu + \varepsilon x = \mu + \varepsilon$, где μ – математическое ожидание случайной величины, а ε – чисто случайная составляющая (остаток).

Необходимо отметить, что оценкой в эконометрике называют как формулу, по которой происходит процесс оценки, так и само значение, полученное по этой формуле.

Все статистические показатели: показатели центральной тенденции, показатели вариации, показатели связи, а также параметры регрессионных уравнений имеют как теоретические, существующие «на самом деле» значения, так и выборочные значения, определяемые на основе ограниченного массива данных и являющихся по сути, оценками теоретических величин:

Обычно теоретические значения обозначаются греческими буквами, а их оценки, полученные в результате эмпирического анализа – соответствующими латинскими.

Для проведения оценки случайных величин используются специальные методы: методы выборочного наблюдения и интерпретации его результатов, метод статистической проверки гипотез.

Существует два *типа оценок – точечные и интервальные*.

Точечная оценка представляет собой конкретное число, которое используется в качестве характеристики случайной величины. Она не дает точного представления о распределении показателя, однако определяет их наиболее вероятные значения. Основные формулы точечной оценки представлены в таблице 3.1

Таблица 3.1 – Основные формулы точечной оценки

№ п/п	Характеристика случайной величины	Обозначение	Формула Оценивания
1.	Ожидаемое значение независимой переменной	μ	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (3.1)
2.	Ожидаемое значение зависимой переменной	y	$\hat{y} = f(x_1, x_2, \dots, x_m)$ $\hat{y} = f(x_1, x_2, \dots, x_m)$ (3.2)
3.	Дисперсия переменной	σ^2	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (3.3)
4.	Случайная ошибка	ε	$e_i = y_i - \hat{y}_i e_i = y_i - \hat{y}_i$ (3.4)
5.	Параметр регрессионной модели	α	$a_i = r_{x_i,y}\frac{S_y}{S_{x_i}} a_i = r_{x_i,y}\frac{S_y}{S_{x_i}}$ * (3.5)

** – формула оценивания параметра регрессионной модели приведена для случая линейной регрессии*

При оценке качества точечных оценок анализируются четыре основных *свойства* оценок:

1. Несмещенность;
2. Эффективность;
3. Состоятельность;
4. Достаточность.

Точечное значение параметра генеральной совокупности называется *несмещенным*, если математическое ожидание значения оценки равно истинному значению параметра, другими словами, если оценки располагаются симметрично относительно истинного значения характеристики генеральной совокупности (рис. 3.1).

Математически это свойство записывается следующим образом: $E(\bar{x}) = \mu$ $E(\bar{x}) = \mu$.

Оценка называется *эффективной*, если она максимально точно описывает истинное значение, то есть мера разброса точечных оценок, получаемых в различных наблюдениях, минимальна (рис. 3.2)

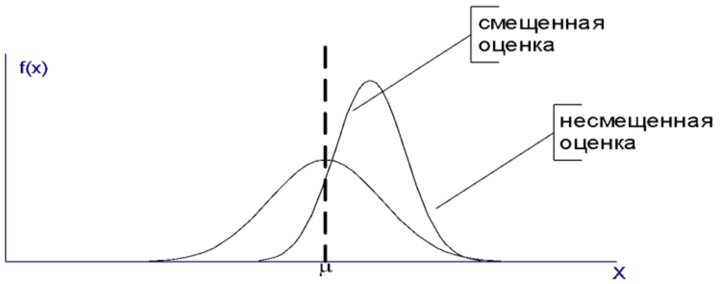


Рис. 3.1 Несмещенность статистической оценки

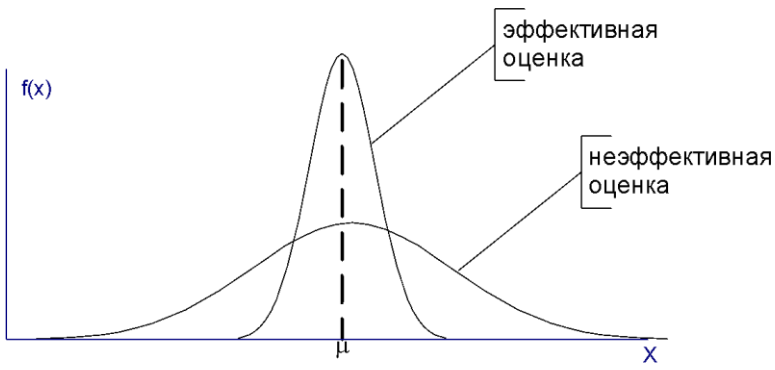


Рис. 3.2 Эффективность статистической оценки

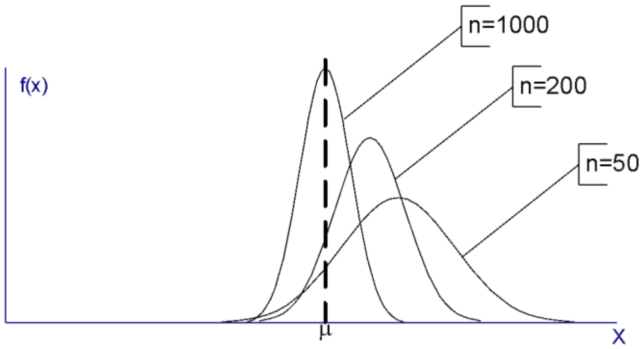


Рис. 3.3 Состоятельность статистической оценки

Часто перед исследователем возникает проблема выбора между эффективной и несмещенной оценками. Выбор в пользу несмещенной оценки делается в том случае, если возможные ошибки не сильно беспокоят исследователя при условии их взаимной компенсации. Если же большие ошибки недопустимы, то выбор будет сделан в пользу эффективной, но смещенной оценки.

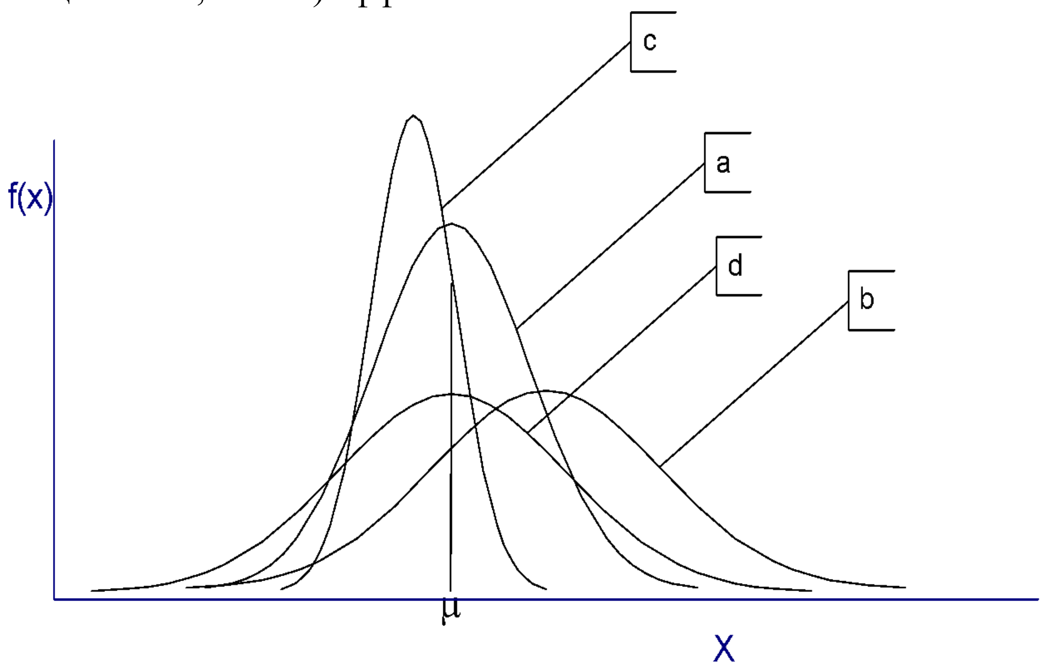
Оценка является *состоятельной*, если по мере увеличения числа единиц в анализируемой выборке ее значение стремится к истинному значению показателя (рис. 3.3).

Под *достаточностью* понимают такое свойство точечной оценки, согласно которому для ее проведения используется максимум информации.

Например, выборочная средняя является наилучшей оценкой генеральной средней, так как соответствует всем четырем свойствам.

Задания

1. Какие из приведенных оценок являются: а) несмещенными; б) эффективными:



2. При помощи трех различных методов проведена оценка средних расходов на питание в вузе. Определите, какая оценка является:

а) несмещенной б) эффективной в) состоятельной,

если истинное значение величины расходов составляет 14 рублей

Число обследованных студентов	1 метод		2 метод		3 метод	
	средние расходы, руб.	СКО	средние расходы, руб.	СКО	средние расходы, руб.	СКО
50	16,0	3,2	13,1	1,0	14	4,0
100	15,4	2,2	13,2	1,0	14,2	3,8
500	13,7	2,0	13,4	0,9	14,1	3,6
1000	14,1	1,8	13,5	0,8	13,8	3,5

Литература: [1, 2, 3, 5]

ЛАБОРАТОРНАЯ РАБОТА 4

Базовые понятия эконометрики

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты.

Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: заключается в том, что большинство величин, с которыми имеет дело исследователь в эконометрике, являются не жестко детерминированными, а вероятностными величинами. Мы можем лишь оценить их значения, предположить с определенной вероятностью границы, в которых они могут находиться на самом деле. Это связано с объективно существующими ошибками наблюдения.

Теоретическая часть

Интервальная оценка представляет собой интервал, в котором с известной вероятностью находится истинное значение исследуемого признака. Такой интервал называется доверительным, а соответствующая ему вероятность – *доверительной вероятностью*. В практическом статистическом анализе большую ценность представляет именно интервальная оценка. Наряду с доверительной вероятностью *(p)* используют термин *уровень значимости* (*α=1-p*).

Рассмотрим общие принципы проведения интервальной оценки. Она проводится в виде определения *доверительных интервалов* – интервалов, в которых с известной вероятностью находится изучаемая переменная. Величина интервала прямо пропорциональна дисперсии рассматриваемой случайной величины и обратно зависима от требуемого уровня значимости. Обычно выбираются уровни значимости 0,1 (10%-й уровень значимости); 0,05 (5%-й уровень значимости); 0,01 (1%-й уровень значимости).

При построении доверительных интервалов для средней величины предполагается, в соответствии с центральной предельной теоремой, что выборочные средние распределяются по нормальному закону распределения с ожидаемым значением, равным генеральной средней. Исходя из этого, доверительные интервалы определяются на основе площади под кривой нормального распределения следующим образом:

$$\mu = \bar{x} \pm u_{a/2} \cdot \frac{\sigma}{\sqrt{n}} \mu = \bar{x} \pm u_{a/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad (4.1)$$

где $u_{\alpha/2}$ – значения, определяемые из таблицы площади под кривой нормального распределения; σ – генеральное среднеквадратическое отклонение; n – число единиц в выборке.

Интервальная оценка используется при анализе коэффициентов регрессии, значений зависимой переменной. Формулы для интервальной оценки представлены в соответствующих темах.

Выделяют следующие основные показатели связи:

1. *Ковариация* – абсолютный показатель связи двух показателей. Характеризует силу и направление линейной связи двух показателей. Различают теоретическую и выборочную ковариацию.

Теоретическая ковариация ($pop.cov(x,y) pop.cov(x,y)$) – это математическое ожидание произведения отклонений двух случайных величин от их средних значений.

Выборочная ковариация рассчитывается по формуле:

$$cov(x,y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) \quad (y_i - \bar{x}) \qquad (4.2)$$

Свойства ковариации:

a. *cov(x,y) = cov(y,x)*
b. *cov(x,x) =* s_x^2
c. если *y = v + w*, то *cov(x,y) = cov(x,v) + cov(x,w)*
d. если *y = a*z*, то *cov(x,y) = a*cov(x,z)*
e. если *a = const*, то *cov(x,a) = 0*

Существует альтернативная формула расчета ковариации:

$$cov(x,y) = \overline{x \cdot y} - \bar{x} \cdot \bar{y} cov(x,y) = \overline{x \cdot y} - \bar{x} \cdot \bar{y} \qquad (4.3)$$

Главным недостатком ковариации как показатели связи является то, что его значение зависит от единиц измерения исходных данных и не имеет критических значений, что затрудняет как сравнение различных совокупностей на предмет силы связи, так и делает невозможным установления критических значений ковариации.

Этот недостаток преодолевает следующий показатель связи – *коэффициент корреляции*. Он является относительным показателем связи и также характеризует силу и направление линейной связи двух признаков, изменяется в пределах от –1 до 1, чем ближе по модулю к единице, тем теснее связь между показателями.

Выделяют теоретический и выборочный коэффициент корреляции. *Теоретический коэффициент корреляции* рассчитывается по формуле:

$$\rho_{x,y} = \frac{pop.cov(x,y)}{\sigma_x \cdot \sigma_y} \rho_{x,y} = \frac{pop.cov(x,y)}{\sigma_x \cdot \sigma_y} \qquad (4.4)$$

где σ_x, σ_y – теоретическое среднеквадратическое отклонение x и y соответственно.

Для расчета *выборочного коэффициента* корреляции используют формулу:

$$r_{x,y} = \frac{cov(x,y)}{s_x \cdot s_y} r_{x,y} = \frac{cov(x,y)}{s_x \cdot s_y} \qquad (4.5)$$

где s_x и s_y – выборочное среднеквадратическое отклонение x и y соответственно.

Вывод о наличии статистически значимой связи можно сделать, оценив значимость коэффициента корреляции. Это можно сделать при помощи *t-критерия* Стьюдента. Для этого проверяется гипотеза о равенстве этого коэффициента 0.

$$t_{\text{расч}} = \sqrt{\frac{r^2}{1-r^2}(n-2)} t_{\text{расч}} = \sqrt{\frac{r^2}{1-r^2}(n-2)} \text{ или} \qquad (4.6)$$

$$t_{\text{расч}} = \sqrt{\frac{r^2}{1-r^2} n} t_{\text{расч}} = \sqrt{\frac{r^2}{1-r^2} n}, \qquad (4.7)$$

где n – число единиц в выборке

Примечание: формула 4.6 применяется при небольших выборках (не более 100 единиц), а формула 4.7 – при больших выборках (более 100 единиц)

Расчетное значение $t_{расч}$ сравнивается с критическим значением при n-2 степенях свободы и требуемом уровне значимости (0,05 или 0,01). Если расчетное значение оказывается больше критического, то делается вывод о том, что коэффициент корреляции значимо отличается от 0 и связь между анализируемыми признаками статистически значима.

Если коэффициент корреляции статистически значим, можно осуществить его *интервальную оценку* с использованием z-распределения Фишера. Расчетное значение z-статистики определяется по формуле:

$$\acute{z} = \frac{1}{2} \cdot ln\frac{1+r}{1-r} \acute{z} = \frac{1}{2} \cdot ln\frac{1+r}{1-r} \qquad (4.8)$$

Доверительный интервал для z' определится следующим образом:

$$z \in \left[\acute{z} \pm Z\alpha_{/2} \sqrt{\frac{1}{n-3}} \right] z \in \left[\acute{z} \pm Z\alpha_{/2} \sqrt{\frac{1}{n-3}} \right] \tag{4.9}$$

где $Z_{\alpha/2}$ – значения нормального распределения для уровня значимости α.

Для того, чтобы определить *доверительный интервал* для самого коэффициента корреляции, необходимо выполнить обратное преобразование полученных границ:

$$r = \frac{e^{2z}+1}{e^{2z}-1} r = \frac{e^{2z}+1}{e^{2z}-1} \tag{4.10}$$

где вместо *z* требуется подставить границы доверительного интервала для z', а *r* покажет значение соответствующей границы для коэффициента корреляции.

В том случае, если оценивается корреляция явления, зависящего более чем от одной независимой переменной, то необходимо определиться с тем, какую связь мы хотим оценить. Если оценивается связь между двумя переменными: результативной и одной объясняющей, то используют коэффициент частной корреляции.

В том случае, если всего используется три переменные, то *коэффициент частной корреляции x* и *y* при условии постоянства *z* рассчитывается по формуле:

$$r_{x,y/z} = \frac{r_{x,y} - r_{x,z} \cdot r_{y,z}}{\sqrt{(1-r_{x,z}^2) \cdot (1-r_{y,z}^2)}} r_{x,y/z} = \frac{r_{x,y} - r_{x,z} \cdot r_{y,z}}{\sqrt{(1-r_{x,z}^2) \cdot (1-r_{y,z}^2)}} \tag{4.11}$$

Если же необходимо установить связь между всеми переменными, то рассчитывается коэффициент полной корреляции:

$$R = \sqrt{\frac{r_{x,y}^2 - 2 r_{x,y} r_{x,z} r_{y,z} + r_{y,z}^2}{1 - r_{x,y}^2}} R = \sqrt{\frac{r_{x,y}^2 - 2 r_{x,y} r_{x,z} r_{y,z} + r_{y,z}^2}{1 - r_{x,y}^2}} \tag{4.12}$$

Для решения конкретных задач в эконометрике используются также и другие показатели связи: коэффициент автокорреляции, коэффициент детерминации и др. Они будут рассмотрены в последующих темах.

Тесты

1. Ковариация является:

a. мерой отклонения индивидуальных значений
b. мерой взаимосвязи двух переменных
c. величиной, всегда положительной
d. абсолютной величиной
e. относительной величиной

Варианты ответа

a. 1,3,4	d. 2,5
b. 1,3,5	e. 2,3,4
c. 2,4	

2. При сравнении коэффициентов ковариации (cov(x,y)) и корреляции ($\rho_{x,y}$) всегда можно сделать вывод, что:

a. cov(x,y) больше $\rho_{x,y}$

b. cov(x,y) меньше $\rho_{x,y}$

c. cov(x,y) равна $\rho_{x,y}$ по модулю и противоположна по знаку

d. cov(x,y) имеет одинаковый знак с $\rho_{x,y}$

3. По какой из формул рассчитывается коэффициент корреляции:

a. $\frac{1}{n}\sum(x_i-\bar{x})^2\frac{1}{n}\sum(x_i-\bar{x})^2$

b. $\frac{1}{n-1}\sum(x_i-\bar{x})(y_i-\bar{y})\frac{1}{n-1}\sum(x_i-\bar{x})(y_i-\bar{y})$

c. $\frac{cov(x,y)}{\sqrt{var(x)\cdot var(y)}}\frac{cov(x,y)}{\sqrt{var(x)\cdot var(y)}}$

4. В чем преимущество коэффициента корреляции перед коэффициентом ковариации

a. дает большую точность расчета

b. обеспечивает сопоставимость анализа по разным выборкам

c. проще методика расчета

d. преимуществ нет

5. Какой показатель связи используется, если требуется оценить связь между двумя факторами при условии постоянства третьего?

a. коэффициент автокорреляции

b. коэффициент частной корреляции

c. коэффициент общей корреляции

d. ковариация

ЛАБОРАТОРНАЯ РАБОТА 5

Парный регрессионный анализ

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты. Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: обуславливается тем, что перечисленные методы могут применяться для «быстрого», поверхностного анализа параметров уравнения регрессии. Полученные на их основе оценки не обладают свойствами несмещенности, эффективности, состоятельности и достаточности, поэтому для серьезного исследования необходимо применять другие методы. Наибольшее распространение получили такие математические методы, как метод наименьших модулей и метод наименьших квадратов.

Теоретическая часть

Парная регрессионная модель (регрессия) – это эконометрическая модель, описывающая зависимость между двумя факторами. Общий вид такой модели:

$$y = f(x) + \varepsilon y = f(x) + \varepsilon \qquad (5.1)$$

Наиболее простой и часто использующейся является *линейная парная регрессионная модель*, имеющая вид:

$$y = \alpha_0 + \alpha_1 x + \varepsilon y = \alpha_0 + \alpha_1 x + \varepsilon \qquad (5.2)$$

Выражение 3.2 представляет собой *спецификацию линейной регрессионной модели*. Вообще под спецификацией модели понимают аналитическое выражение описывающей модель функции.

Само *уравнение линейной регрессии* имеет вид:

$$\hat{y} = f(x) = \alpha_0 + \alpha_1 x\, \hat{y} = f(x) = \alpha_0 + \alpha_1 x \qquad (5.3)$$

где a_0 и a_1 – оценки теоретических коэффициентов регрессии α_0 и α_1

Следовательно, регрессионную модель можно представить в виде: $y = y + \varepsilon y = y + \varepsilon$, где $\hat{y}$ – объясненная на основе построенной модели составляющая y, а ε – чисто случайная составляющая.

Основная задача регрессионного анализа после спецификации модели – оценка неизвестных параметров – α_0 и α_1, дающих наибольшее приближение модели к эмпирическим данным:

$$y_i = \alpha_0 + \alpha_1 x_i + e_i\; y_i = \alpha_0 + \alpha_1 x_i + e_i \qquad (5.4)$$

где a_0 и a_1 – оценки неизвестных параметров; e – оценка случайной компоненты.

Для определения коэффициентов можно использовать различные методы (рис. 5.1):

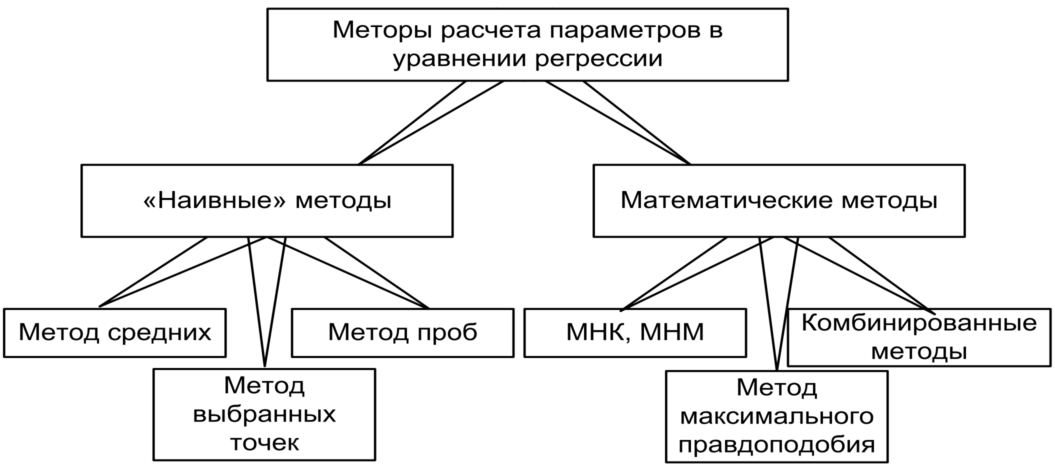


Рис. 5.1 Основные методы расчета коэффициентов регрессии

Метод средних применяется в том случае, когда в уравнении регрессии присутствует только один неизвестный параметр (например, a_1 – $y = a_1x$). В этом случае его значение определится следующим образом:

$$a_1 = \frac{\bar{y}}{\bar{x}} a_1 = \frac{\bar{y}}{\bar{x}} \qquad (5.5)$$

Метод проб заключается в том, что всем параметрам, кроме одного, задаются фиксированные значения, исходя из особенностей эмпирических данных. Значение последнего, неизвестного параметра определяется по методу средних. Например, если зафиксировать значение, принимаемое y при x, равном 0 (нулевой уровень $y – y_0$), то параметр a_1 определится по формуле:

$$a_1 = \frac{\bar{y}-y_0}{\bar{x}} a_1 = \frac{\bar{y}-y_0}{\bar{x}} \qquad (5.6)$$

Эта процедура может повторяться (для различных зафиксированных значений) до тех пор, пока качество теоретической модели не станет удовлетворительным.

Метод выбранных точек основан на визуальном анализе корреляционного поля и выборе точек, наиболее точно отражающих тенденции развития анализируемого явления. Количество точек должно совпадать с количеством неизвестных параметров. Так, для парной линейной регрессии выбирается 2 точки. Через них проходит только одна прямая, уравнение которой определится как решение системы относительно a_0 и a_1:

$$\begin{cases} y_1 = \alpha_0 + \alpha_1 x_1 \\ y_2 = \alpha_0 + \alpha_1 x_2 \end{cases} \begin{cases} y_1 = \alpha_0 + \alpha_1 x_1 \\ y_2 = \alpha_0 + \alpha_1 x_2 \end{cases},$$

где $(x_1;y_1)$ и $(x_2;y_2)$ – координаты выбранных точек.

В результате решения такой системы можно рассчитать значения a_0 и a_1 по формулам:

$$a_0 = \frac{y_1x_2 - y_2x_1}{x_2 - x_1} \qquad (5.7)$$

$$a_1 = \frac{y_2 - y_1}{x_2 - x_1} \qquad (5.8)$$

Перечисленные методы могут применяться для «быстрого», поверхностного анализа параметров уравнения регрессии. Полученные на их основе оценки не обладают свойствами несмещенности, эффективности, состоятельности и достаточности, поэтому для серьезного исследования необходимо применять другие методы. Наибольшее распространение получили такие математические методы, как метод наименьших модулей и метод наименьших квадратов. Существуют также методы, совмещающие достоинства этих методов, и преодолевающие их недостатки (в частности, *функция Хубера*).

В общем виде смысл математических методов можно определить как решение задачи минимизации функционала F, формируемого на основе суммирования отклонений эмпирических данных от результата расчета (чисто случайных составляющих) по регрессионной модели:

$$F = \sum_{i=1}^{n} g(y_i - \bar{y}_i) = \sum_{i=1}^{n} g(e_i)$$

(5.9)

где *g()* – функция, определяющая аналитическую форму измерения разброса фактических данных от модели.

Наиболее распространены два вида функции *g():*

$$g(y_i - \hat{y}_i) = (y_i - \hat{y}_i) g(y_i - \hat{y}_i) = (y_i - \hat{y}_i), \quad (5.10)$$

$$g(y_i - \hat{y}_i) = |y_i - \hat{y}_i| g(y_i - \hat{y}_i) = |y_i - \hat{y}_i| \quad (5.11)$$

в том случае, если расчет выполнен по формуле (5.10), говорят о *методе наименьших квадратов* (МНК), если по формуле (5.11) – о *методе наименьших модулей* (МНМ).

Результаты сравнительного анализа этих методов приведены ниже:

	метод наименьших квадратов	метод наименьших модулей
достоинства	**легкость вычислительной процедуры, хорошие статистические свойства**	**нечуствительность к «выбросам» (робастность)**
недостатки	**чувствительность к «выбросам»**	**сложность вычислительной процедуры, неоднозначность, равнозначность для метода больших и маленьких отклонений при условии равенства их общей суммы**

Рис. 5.2 Сравнительный анализ основных математических методов определения параметров уравнения регрессии.

Для совмещения достоинств этих методов разработана более сложная кусочно заданная *функция Хубера* :

$$g(y_i - \hat{y}_i) = \begin{cases} (y_i - \hat{y}_i)^2, |y_i - \hat{y}_i| < c \\ 2c \cdot (y_i - \hat{y}_i) - c^2, (y_i - \hat{y}_i) \geq \tilde{n} \\ -2c \cdot (y_i - \hat{y}_i) - c^2, (y_i - \hat{y}_i) \leq -\tilde{n} \end{cases}$$

$$g(y_i - \hat{y}_i) = \begin{cases} (y_i - \hat{y}_i)^2, |y_i - \hat{y}_i| < c \\ 2c \cdot (y_i - \hat{y}_i) - c^2, (y_i - \hat{y}_i) \geq \tilde{n} \\ -2c \cdot (y_i - \hat{y}_i) - c^2, (y_i - \hat{y}_i) \leq -\tilde{n} \end{cases} \quad (5.12)$$

где с – параметр, показывающий границу, начиная с которой в качестве меры отклонения используется модуль (при меньших – квадрат), чем он больше, тем сильнее чувствительность .

Для снижения чувствительности *g()* к выбросам (значениям, выбивающимся из общей тенденции) Пиндайк и Рубинфелд ввели функцию:

$$g(y_i - \hat{y}_i) = \begin{cases} (y_i - \hat{y}_i)^2, |y_i - \hat{y}_i| < c \\ 0, |y_i - \hat{y}_i| \geq \tilde{n} \end{cases} g(y_i - \hat{y}_i) = \begin{cases} (y_i - \hat{y}_i)^2, |y_i - \hat{y}_i| < c \\ 0, |y_i - \hat{y}_i| \geq \tilde{n} \end{cases}$$

(5.13)

Наибольшее распространение в настоящее время получил метод наименьших квадратов. Рассмотрим условия применения МНК, основной алгоритм расчета параметров и свойства оценок параметров, полученных в результате его применения.

Условия применения (предпосылки) МНК (условия Гаусса – Маркова).

1. Математическое ожидание случайного отклонения равно 0 для всех наблюдений: $M(\varepsilon)=0$

2. Дисперсия случайных отклонений постоянна: $s_\varepsilon^2 = const s_\varepsilon^2 = const$
3. Случайные отклонения независимы друг от друга:
$r_{e_i e_j} = 0\ \forall i,j$
4. Случайное отклонение независимо от объясняющих переменных регрессионной модели: $r_{e,x} = 0 r_{e,x} = 0$
5. Модель линейна относительно параметров

Математически условие минимизации квадратов отклонений запишется следующим образом:

$$F = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 \cdot x_i)^2 \to min$$

(5.14)

найдем минимум F относительно a_0 и a_1, вычислив частные производные F по a_0 и a_1 и прировняв их к 0:

$$\begin{cases} \dfrac{\partial F}{\partial a_0} = -2 \cdot \sum_{i=1}^{n} (y_i - a_0 - a_1 \cdot x_i) = 0 \\ \dfrac{\partial F}{\partial a_1} = -2 \cdot \sum_{i=1}^{n} x_i (y_i - a_0 - a_1 \cdot x_i) = 0 \end{cases}$$

(5.15)

преобразуем полученную систему:

$$\begin{cases} \sum_{i=1}^{n}(y_i - a_0 - a_1 \cdot x_i) = 0 \\ \sum_{i=1}^{n} x_i(y_i - a_0 - a_1 \cdot x_i) = 0 \end{cases}$$

(5.16)

раскрыв скобки, получаем *стандартную форму нормальных уравнений* для вычисления коэффициентов линии регрессии.

$$\begin{cases} a_0 \cdot n + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a_0 \cdot \sum_{i=1}^{n} x_i + a_1 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \cdot x_i \end{cases}$$

(5.17)

решив это уравнение относительно a_0 и a_1 , получим оценки параметров теоретического уравнения α_0 и α_1, обладающие всеми основными свойствами качественных оценок: (по теореме Гаусса-Маркова): несмещенность, эффективность, состоятельность. Это обуславливает широкое использование МНК в эконометрических расчетах.

Для линейной модели существует упрощенный способ расчета параметров, основанный на решении системы (3.14)

$$\begin{cases} a_1 = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \dfrac{cov(x,y)}{s_x^2} = r_{x,y} \dfrac{s_y}{s_x} \\ a_0 = \bar{y} - a_1 \cdot \bar{x} \end{cases}$$

(5.18)

Однако не все фактические данные могут быть описаны при помощи линейной модели. В этом случае используется нелинейная регрессия. Для того, чтобы определить, какую аналитическую форму регрессионной модели выбрать, используют следующий алгоритм (применяется для каждой гипотезы об определенной аналитической форме регрессии) (рис. 5.3). Следует отметить, что этот алгоритм применим только в случае монотонной зависимости между факторами.

Задание 1

1. Запишите спецификацию линейной регрессии зависимости экспорта от импорта. По представленным в таблице данным о внешнеэкономической деятельности РФ определите значения параметров модели, используя различные методы.

	Экспорт, млн. долл.	Импорт, млн. долл.
январь	6655	3682
февраль	6593	4004
март	8341	4662

	Экспорт, млн. долл.	Импорт, млн. долл.
апрель	9407	5131
май	8495	4683
июнь	8236	4955
июль	9215	5459
август	9868	5137
сентябрь	9736	5129
октябрь	10077	5876
ноябрь	9457	5744
декабрь	11170	6505

Для каждой модели рассчитайте значения коэффициента детерминации. По результатам расчета заполните таблицу:

№$_{п/п}$	Метод расчета	Значение параметра a_0	Значение параметра a_1	Коэффициент детерминации R^2
1.	…	…	…	…

2. Проведите линеаризацию следующих функций:

№ п/п	Вид модели	Аналитическое выражение	Замена переменных	Линеаризованное уравнение
1.	…	$y = e^{a_0 + a_1 x + \varepsilon}$	…	…
2.	…	$y = a_0 + \frac{a_1}{x^2} + \varepsilon$ $y = a_0 + \frac{a_1}{x^2} + \varepsilon$	…	…
3.	…	$y = a_0 + a_1\sqrt{x} + \varepsilon$ $y = a_0 + a_1\sqrt{x} + \varepsilon$	…	…
4.	…	$y = a_0 + a_1\cos(x) + \varepsilon$ $y = a_0 + a_1\cos(x) + \varepsilon$		

3. По данным из задания 1 постройте 3 нелинейные парные регрессии. Обоснуйте выбор моделей при помощи соответствующего алгоритма. Оцените качество моделей по коэффициенту детерминации и сравните с моделью линейной регрессии.

4. Оцените статистическую значимость коэффициентов уравнения линейной регрессии, построенной в задании 1, при уровне значимости 0,05; 0,01.

Литература: [1, 2, 5, 6].

ЛАБОРАТОРНАЯ РАБОТА 6

Парный регрессионный анализ

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты.

Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: обуславливается тем, что перечисленные методы могут применяться для «быстрого», поверхностного анализа параметров уравнения регрессии.

Теоретическая часть

Расчет параметров *нелинейных регрессионных моделей* основан на том же методе, что и для линейной регрессии. Основное требование – уравнение регрессии должно быть либо линейно относительно параметров, либо преобразуемо в такое уравнение (это преобразование называется *линеаризацией*). В случае параболической модели и полиномиальной модели более высокой степени система уравнений для определения параметров претерпевает очевидные изменения (6.1):

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i \cdot x_i \\ a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} y_i \cdot x_i^2 \end{cases} \tag{6.1}$$

где a_0, a_1 и a_2 – оценки параметров в уравнении регрессии $y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \varepsilon$

Упорядочивание исходных данных по величине x (объясняющей переменной)

Расчет и по фактическим данным (табл. 6.1):

Таблица 6.1

Определение :

значение y, соответствующее (если среди исходных данных существует)

по формуле: , (6.2)

где , – фактические значения x, между которыми заключено значение , а y_i и y_{i+1} – соответствующие им значения зависимой переменной

Расчет разницы и выбор формы, которой соответствует наименьшая разница

Рис. 6.1 Алгоритм определения аналитической формы регрессии

Рассмотренные модели также можно комбинировать, получая линейные преобразования для других исходных моделей.

В случае линеаризации происходит замена переменных в уравнении регрессии с тем, чтобы привести его к линейному виду. Обратите внимание, что линеаризованы могут быть функции с числом параметров, равным числу параметров в соответствующей линейной модели (для парной регрессии – с двумя параметрами), поэтому по сравнению с таблицей 6.1 аналитические выражения для основных функций претерпели определенные изменения.

Таблица 6.2 – Линеаризация основных видов регрессионных моделей

№ п/п	Вид модели	Аналитическое выражение	Замена переменных	Линеаризованное уравнение
1.	Степенная	$y = \alpha_1 x^{\alpha_2} \cdot \varepsilon$ $lny = ln\alpha_1 + \alpha_2 lnx + \varepsilon'$	$x' = lnx$ $y' = lny$	$y' = ln\alpha_1 + \alpha_2 \cdot x' + \varepsilon'$
2.	Показательная	$y = \alpha_1 \cdot \alpha_2^x \cdot \varepsilon$ $lny = ln\alpha_1 + \alpha_2 x + \varepsilon'$	$y' = lny$ $\alpha_1{}' = ln\alpha_1$	$y' = \alpha'_1 + \alpha_2 \cdot x + \varepsilon'$
3.	Логарифмическая	$y = \alpha_0 + \alpha_1 \cdot lnx + \varepsilon$	$x' = lnx$	$y = \alpha_1 + \alpha_2 \cdot x' + \varepsilon$
4а.	Гиперболическая тип 1	$y = \alpha_0 + \frac{\alpha_1}{x} + \varepsilon$	$x' = {}^1/_x$	$y = \alpha_1 + \alpha_2 \cdot x' + \varepsilon$
4б.	Гиперболическая тип 2	$y = \frac{1}{\alpha_0 + \alpha_1 \cdot x + \varepsilon}$,	$y' = {}^1/_y$	$y' = \alpha_1 + \alpha_2 \cdot x + \varepsilon$
4в.	Гиперболическая тип 3	$y = \frac{x}{\alpha_0 + \alpha_1 \cdot x + \varepsilon}$	$x' = {}^1/_x$ $y' = {}^1/_y$	$y' = \alpha_1 + \alpha_2 \cdot x' + \varepsilon$
5.	Логистическая	$y = \frac{1}{\alpha_0 + \alpha_1 \cdot e^{-x} + \varepsilon}$	$x' = e^{-x}$ $y' = {}^1/_y$	$y' = \alpha_1 + \alpha_2 \cdot x' + \varepsilon$
8.	Тригонометрическая	$y = \alpha_0 + \alpha_1 \cdot \sin(x) + \varepsilon$	$x' = \sin(x)$	$y = \alpha_1 + \alpha_2 \cdot x' + \varepsilon$

В оценке качества парных регрессионных моделей можно выделить следующие основные этапы (рис. 6.2):

1. Анализ адекватности модели в целом
2. Анализ точности определения оценок коэффициентов регрессии
3. Проверка статистической значимости коэффициентов регрессионного уравнения
4. Интервальная оценка коэффициентов регрессионного уравнения при заданном уровне значимости
5. Определение доверительных интервалов для зависимой переменной (для среднего значения и для индивидуальных значений)

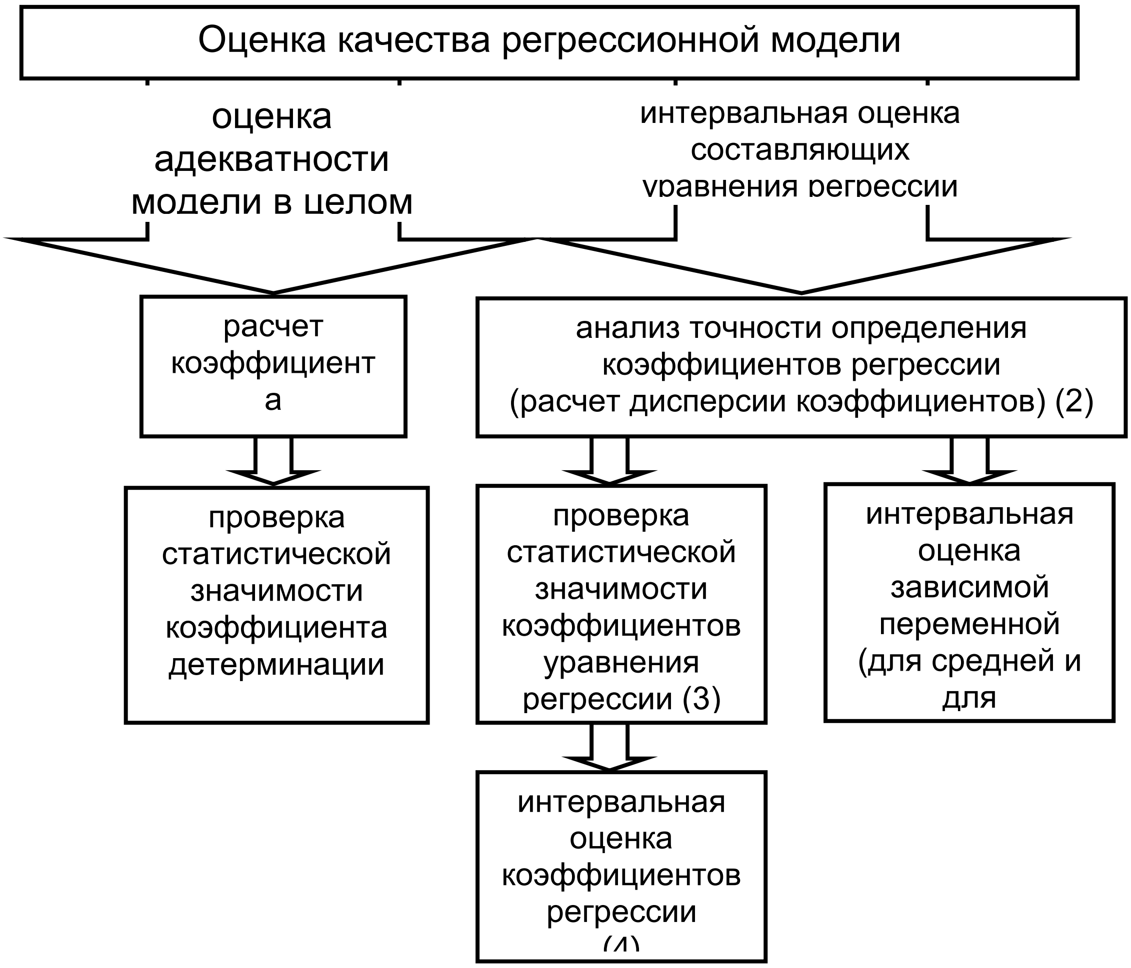


Рис. 6.2 Анализ качества регрессионной модели

Рассмотрим каждый этап подробно:

I. Для определения адекватности модели в целом используется *теоретический коэффициент детерминации (коэффициент детерминации, R^2).*

Коэффициент детерминации показывает, какая доля вариации независимой переменной объяснена на основе построенной регрессионной модели:

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} R^2 = \frac{s_{\hat{y}}^2}{s_y^2} \tag{6.3}$$

где $s_y^2 s_y^2$ – фактическая дисперсия зависимой переменной, $s_{\hat{y}}^2 s_{\hat{y}}^2$ – дисперсия оценочных значений зависимой переменной, полученных на основании модели:

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{6.4}$$

$$s_{\hat{y}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{6.5}$$

R^2 принимает значения от 0 до 1, причем чем ближе его значение к 1, тем лучше построенная модель описывает фактическую зависимость.

Так как $y = \hat{y} + e y = \hat{y} + e$, общая дисперсия $s_y^2 s_y^2$ может быть представлена как сумма дисперсии объясненной составляющей $s_{\hat{y}}^2 s_{\hat{y}}^2$ и дисперсии случайного остатка $s_e^2 s_e^2$:

$$s_y^2 = s_{\hat{y}}^2 + s_e^2 s_y^2 = s_{\hat{y}}^2 + s_e^2 \qquad (6.6)$$

где

$$s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 \qquad (6.7)$$

В этом случае формула (3.21) будет иметь вид:

$$R^2 = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2} R^2 = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2} \qquad (6.8)$$

откуда можно сделать вывод, что качество модели будет тем выше, чем меньше вариация случайного остатка.

После несложных преобразований получаем формулы, часто используемые в практических расчетах:

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (6.9)$$

Значение теоретического коэффициента детерминации связано со значением линейного коэффициента корреляции между теоретическими и фактическими значениями результативного признака:

$$R^2 = r_{y\hat{y}}^2 R^2 = r_{y\hat{y}}^2 \qquad (6.10)$$

Вывод о приемлемости регрессионной модели для описания фактических данных можно сделать, учитывая объем анализируемой совокупности, число переменных и прочие факторы. Обычно значение R^2 не превышает 0,7 (исключение составляют временные ряды с четко выраженным трендом, для которых это значение приближается к 1). Для точной оценки статистической значимости коэффициента детерминации используют F-критерий:

$$F = \frac{R^2}{1-R^2} \cdot (n-2) F = \frac{R^2}{1-R^2} \cdot (n-2) \qquad (6.11)$$

Полученное фактическое значение сравнивают с критическим $F_{\alpha;\,1;\,n-2}$, если расчетное значение оказывается больше критического, то нулевая гипотеза отвергается и делается вывод о статистической значимости коэффициента детерминации.

Для статистически значимого коэффициента детерминации может быть проведена интервальная оценка при помощи z-распределения Фишера. Порядок оценки следующий: рассчитывают величину $R = \sqrt{R^2} R = \sqrt{R^2}$ и для нее проводят оценку по тому же алгоритму, что и для коэффициента корреляции (см. тему 2).

Затем полученные значения доверительных пределов возводятся в квадрат и получаем искомую интервальную оценку.

II. Анализ точности определения оценок регрессии осуществляется путем вычисления *дисперсий коэффициентов регрессии*. Для линейной регрессионной модели $y = a_0 + a_1 x + ey = a_0 + a_1 x + e$ значения выборочных дисперсий будут равны:

$$s_{a_1}^2 = \frac{s_e^2}{\sum(x_i - \bar{x})^2} s_{a_1}^2 = \frac{s_e^2}{\sum(x_i - \bar{x})^2} \qquad (6.12)$$

$$s_{a_0}^2 = \frac{s_e^2 \cdot \sum x_i^2}{n \cdot \sum(x_i - \bar{x})^2} = \overline{x^2} \cdot s_{a_1}^2 s_{a_0}^2 = \frac{s_e^2 \cdot \sum x_i^2}{n \cdot \sum(x_i - \bar{x})^2} = \overline{x^2} \cdot s_{a_1}^2 \qquad (6.13)$$

Таким образом, оценки коэффициентов будут тем точнее, чем меньше значение необъясненной дисперсии.

Значения дисперсий коэффициентов регрессии и корни квадратные из них – *стандартные ошибки коэффициентов регрессии* – используются на следующем этапе для проверки статистической значимости коэффициентов регрессии.

III. Оценка статистической значимости коэффициента регрессии осуществляется путем проверки *гипотезы о равенстве этого коэффициента 0*. Для коэффициента a_1 такая гипотеза будет иметь вид:

$H_0 : a_1 = 0$

$H_1 : a_1 \neq 0$

Для проверки этой гипотезы пользуются *t-статистикой*:

$$t = \frac{a_1}{s_{a_1}} t = \frac{a_1}{s_{a_1}} \qquad (6.14)$$

это соотношение имеет распределение Стьюдента с числом степеней свободы, равным $(n-2)$. Расчетное значение t сравнивают с критическим $t_{\text{ст}} = t_{\frac{\alpha}{2};n-2} t_{\text{ст}} = t_{\frac{\alpha}{2};n-2}$, взятым из таблицы где α – уровень значимости. Если фактическое значение оказывается *больше* критического, то нулевая гипотеза отвергается, и делается вывод о статистической значимости коэффициента регрессии. В противном случае считается, его значением можно пренебречь, и рассматривать модель с меньшим числом параметров.

Для предварительной «грубой» оценки статистической значимости коэффициентов регрессии можно пользоваться следующим правилом:

Таблица 6.3

Правило «грубой» оценки статистической значимости коэффициентов регрессионного уравнения

№п/п	Значения t	Описание значимости коэффициента	Доверительная вероятность
1	$\lvert t \rvert \leq 1$	практически незначим	меньше 0,7
2	$1 < \lvert t \rvert \leq 2$	относительно (слабо) значим	от 0,7 до 0,95
3	$2 < \lvert t \rvert \leq 3$	существенно значим	от 0,95 до 0,99

4	$3 < \lvert t \rvert$	гарантированно значим	больше 0,99

Это правило позволяет достаточно точно установить значимость коэффициентов регрессии при *n > 10*.

IV. Интервальная оценка коэффициентов регрессионного уравнения осуществляется для того, чтобы получить более полное представление о характере регрессионной зависимости между переменными. Ее результатом будут доверительные интервалы для каждого коэффициента:

$$\left(a_0 - t_{\frac{\alpha}{2};n-2} \cdot s_{a_0}; a_0 + t_{\frac{\alpha}{2};n-2} \cdot s_{a_0}\right)$$

для α_0 –

$$\left(a_0 - t_{\frac{\alpha}{2};n-2} \cdot s_{a_0}; a_0 + t_{\frac{\alpha}{2};n-2} \cdot s_{a_0}\right) \tag{6.15}$$

$$\left(a_1 - t_{\frac{\alpha}{2};n-2} \cdot s_{a_1}; a_1 + t_{\frac{\alpha}{2};n-2} \cdot s_{a_1}\right)$$

для α_1 –

$$\left(a_1 - t_{\frac{\alpha}{2};n-2} \cdot s_{a_1}; a_1 + t_{\frac{\alpha}{2};n-2} \cdot s_{a_1}\right) \tag{6.16}$$

Доверительный интервал определяет границы, в которых будет находиться значение теоретического коэффициента регрессии с уровнем значимости α.

Уровень значимости α определяется исходя из требуемой точности. Обычно – 0.1, 0.05 или 0.01.

V. Расчет *доверительных интервалов для зависимой переменной* позволяет решить две задачи: во-первых, провести интервальную оценку математического ожидания зависимой переменной для конкретного значения независимой переменной и заданного уровня значимости, и, во-вторых, определить границы, за пределами которых может оказаться не более чем α-ая доля индивидуальных значений зависимой переменной для конкретного значения независимой переменной.

Первая задача решается путем нахождения доверительного интервала для зависимой переменной по формуле:

$$\left(a_0 + a_1 x_p - t_{\frac{\alpha}{2};n-2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum (x_i - \bar{x})^2}}; a_0 + a_1 x_p + t_{\frac{\alpha}{2};n-2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum (x_i - \bar{x})^2}};\right)$$

$$\left(a_0 + a_1 x_p - t_{\frac{\alpha}{2};n-2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum (x_i - \bar{x})^2}}; a_0 + a_1 x_p + t_{\frac{\alpha}{2};n-2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum (x_i - \bar{x})^2}};\right)$$

(6.17)

Для каждого значения x_p из области, в которой находятся значения независимой переменной, определяются доверительные интервалы. Они будут наименьшими при $x_p = \bar{x} x_p = \bar{x}$ и увеличиваться по мере удаления от среднего значения (рис. 6.3).

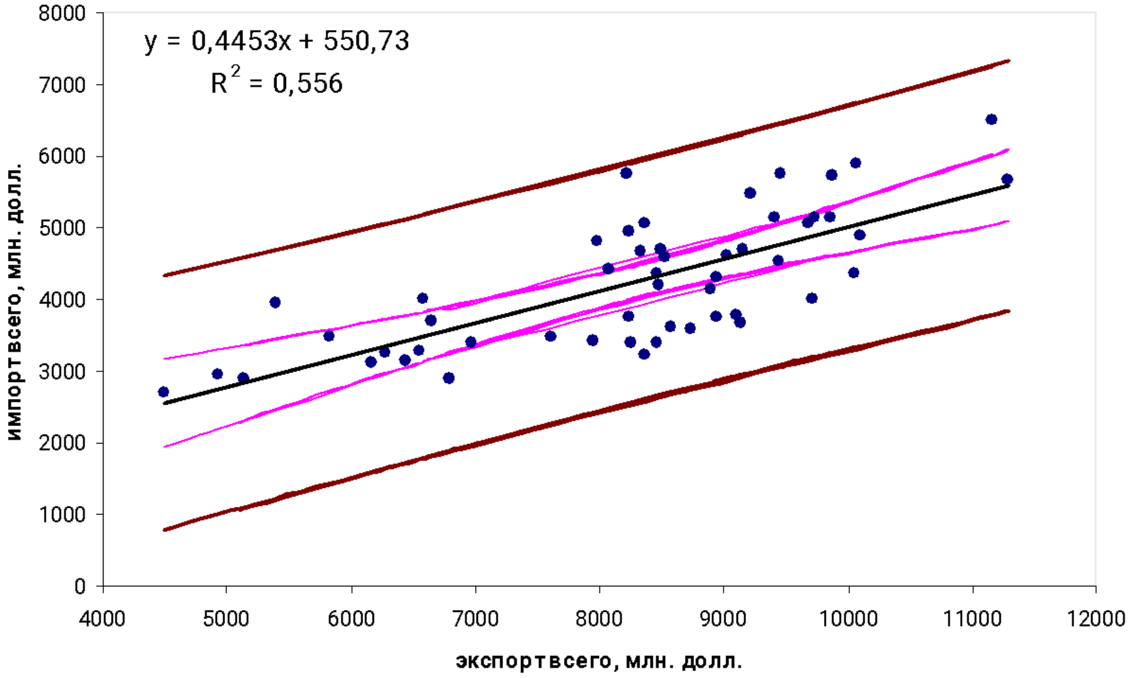


Рис. 6.3 Доверительные интервалы для зависимой переменной. Более широкие интервалы – для индивидуальных значений, более узкие – для средних (уровень значимости 1 %). По данным Госкомстата за 1999 – 2003 год

Вторая задача решается путем вычисления доверительного интервала

$$\left(a_0 + a_1 x_p - t_{\frac{\alpha}{2};n-2}\cdot s_e\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-x_p)^2}{\sum(x_i-\bar{x})^2}};\, a_0 + a_1 x_p + t_{\frac{\alpha}{2};n-2}\cdot s_e\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-x_p)^2}{\sum(x_i-\bar{x})^2}};\right)$$

$$\left(a_0 + a_1 x_p - t_{\frac{\alpha}{2};n-2}\cdot s_e\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-x_p)^2}{\sum(x_i-\bar{x})^2}};\, a_0 + a_1 x_p + t_{\frac{\alpha}{2};n-2}\cdot s_e\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-x_p)^2}{\sum(x_i-\bar{x})^2}};\right) \quad (6.18)$$

Как видно, во втором случае доверительные пределы будут шире, что свидетельствует о том, что оценка индивидуальных значений осуществляется с меньшей точностью. (см. рис. 6.3)

Регрессионный анализ является эффективным инструментом познания экономической действительности, однако существуют *ограничения*, нарушение которых может привести к неверным выводам и некачественной трактовке результатов. Эти ограничения связаны со следующими ошибками:

1. Использование регрессионной модели *для прогнозирования вне границ изменения наблюдаемых данных.* Прогнозирование на основе регрессионных моделей может осуществляться только на основе экстраполяции, в противном случае возможны серьезные ошибки.

2. *Смешение понятий причинно-следственной и регрессионной зависимости*. По наличию статистической связи нельзя делать вывод о том, что взаимосвязанные явления влияют друг на друга.

3. *Перенесение прошлых тенденций в ряде динамики на будущее*. Поскольку исторические условия в прошлом и будущем различаются.

4. *Выявление нереальных (ошибочных) связей.* Для проведения

регрессионного анализа и трактовки его результатов необходима теоретическая гипотеза о взаимосвязи исследуемых переменных.

В том случае, если удастся избежать перечисленных выше ошибок, результаты регрессионного анализа могут с успехом использоваться при выявлении экономических закономерностей, социально-экономическом прогнозировании и разработке экономической политики.

Тесты

1. Что из перечисленного относится к ограничениям регрессионного анализа?

1) возможность прогнозирования только внутри границ измеряемых данных

2) возможность учета не более двух факторов

3) вероятность получения нереальных связей

4) сложности расчета значений параметров

Варианты ответа

a. 1,2,3,4

b. 1,2

c. 2,3

d. 1,4

e. 2,4

2. Согласно правила «грубой» оценки статистической значимости коэффициентов регрессионного уравнения, какое из значений t-статистики свидетельствует о существенной значимости коэффициента?

a. –0,2

b. 0,91

c. 1,34

d. 2,12

e. 7,22

3. Какую функциональную форму модели следует выбрать, если известно, что:

Модель	$\bar{y}_s$	$\bar{\bar{y}}_s$
линейная	100	300
показательная	75	300
гиперболическая	100	130
логарифмическая	100	170

a. линейная

b. гиперболическая

c. показательная

d. логарифмическая

4. Значим ли коэффициент детерминации, если его значение – 0,60 а n=22 ($F_{0,05;2;20}$ = 3,49, $F_{0,01;2;20}$ = 5,85)

a. значим на 5 %-м уровне

b. значим на 1 %-м уровне

c. не значим
d. недостаточно данных для ответа

5. Какой из доверительных интервалов для зависимой переменной шире: для индивидуального значения или для среднего?

a. для среднего
b. для индивидуальных значений
c. нельзя сказать однозначно

Литература: [1, 2, 5].

ЛАБОРАТОРНАЯ РАБОТА 7

Множественный регрессионный анализ

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты.

Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: объясняется тем, что для множественной регрессионной модели актуален вопрос о том, какова сила влияния различных факторов на значение зависимой переменной. Для этого используются два основных метода. Первый основан на построении регрессионной модели в стандартизированной форме, второй - на расчете частных коэффициентов эластичности.

Теоретическая часть

Множественная регрессионная модель (множественная регрессия) представляет собой модель, связывающую несколько независимых (объясняющих) переменных с одной результативной. Общий вид модели, включающей m независимых переменных:

$$y = f(x_1, x_2, \dots x_m) + \varepsilon y = f(x_1, x_2, \dots x_m) + \varepsilon \qquad (7.1)$$

Обычно рассматривают линейную модель, поскольку, как было показано в предыдущей теме, большинство нелинейных моделей достаточно легко сводятся к линейной путем линеаризации. *Спецификация модели линейной множественной регрессии* имеет вид:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m + \varepsilon = \alpha_0 + \sum_{i=1}^{m} \alpha_i x_i + \varepsilon \quad (7.2)$$

Уравнение линейной множественной регрессии будет выглядеть следующим образом:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m = a_0 + \sum_{i=1}^{m} a_i x_i \quad (7.3)$$

В случае применения множественного регрессионного анализа особое внимание следует уделить вопросам отбора переменных для анализа. Для этого используются два подхода: «сверху вниз» и «снизу вверх». Достоинства и недостатки каждого подхода были рассмотрены в теме 1, поэтому мы сразу затронем вопрос практической их реализации. На первом этапе построения модели составляется *матрица корреляции* размером *(m+1) x (m+1)*, где *m* – общее число всех возможных независимых переменных (факторов). В нее помещаются коэффициенты корреляции между факторами и результативным признаком, а также попарно между всеми факторами. В ячейке r_{ij} указывается коэффициент корреляции между *i*-м и *j*-м фактором. Эта матрица будет симметричной относительно главной диагонали, причем на диагонали будут значения, равные 1:

$$\begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} \cdots & r_{x_1x_m} & r_{x_1y} \\ r_{x_2x_1} & r_{x_2x_2} \cdots & r_{x_2x_m} & r_{x_2y} \\ \cdots & \cdots & \cdots & \cdots \\ r_{x_mx_1} & r_{x_mx_2} \cdots & r_{x_mx_m} & r_{x_my} \\ r_{yx_1} & r_{yx_2} \cdots & r_{yx_m} & r_{yy} \end{pmatrix} = \begin{pmatrix} 1 & r_{x_1x_2} \cdots & r_{x_1x_m} & r_{x_1y} \\ r_{x_2x_1} & 1 \cdots & r_{x_2x_m} & r_{x_2y} \\ \cdots & \cdots & \cdots & \cdots \\ r_{x_mx_1} & r_{x_mx_2} \cdots & 1 & r_{x_my} \\ r_{yx_1} & r_{yx_2} \cdots & r_{yx_m} & 1 \end{pmatrix} \quad (7.4)$$

После этого в модель включаются факторы, для которых значение r_{ij} по модулю больше заданного критического значения. Обычно критическое значение устанавливают на уровнях 0.5 – 0.9. В то же время не следует включать в модель переменные, между которыми наблюдается тесная взаимосвязь (высокие значения r_{ij} в соответствующей ячейке). невыполнение этого условия может привести к некорректному построению модели (см. *мультиколлинеарность*).

По аналогии с парной регрессией для определения параметров множественной регрессии могут использоваться различные методы, однако чаще всего применяется МНК. Для его применения необходимо выполнение следующих *предпосылок*:

1. Математическое ожидание случайного отклонения равно 0 для всех наблюдений *M(ε)=0*
2. Дисперсия случайных отклонений постоянна $s_{\varepsilon}^2 = const$
3. Случайные отклонения независимы друг от друга $r_{e_ie_j} = 0 \quad \forall i, j$

4. Случайное отклонение независимо от объясняющих переменных регрессионной модели: $r_{e\,x_i} = 0 \quad \forall i$

5. Модель линейна относительно параметров

6. Отсутствие сильной взаимозависимости между объясняющими переменными

7. Случайные ошибки имеют нормальное распределение

При выполнении перечисленных предпосылок неизвестные коэффициенты уравнения регрессии $a_1, a_2, \ldots a_m$ найдутся из решения системы уравнений:

$$\begin{cases} na_0 + a_1\sum_{i=1}^{n} x_{1i} + a_2\sum_{i=1}^{n} x_{2i} \ldots + a_m\sum_{i=1}^{n} x_{mi} = \sum_{i=1}^{n} y_i \\ a_0\sum_{i=1}^{n} x_{1i} + a_1\sum_{i=1}^{n} x_{1i}^2 + a_2\sum_{i=1}^{n} x_{1i}\cdot x_{2i} \ldots + a_m\sum_{i=1}^{n} x_{1i}\cdot x_{mi} = \sum_{i=1}^{n} y_i\cdot x_{1i} \\ a_0\sum_{i=1}^{n} x_{2i} + a_1\sum_{i=1}^{n} x_{1i}\cdot x_{2i} + a_2\sum_{i=1}^{n} x_{2i}^2 \ldots + a_m\sum_{i=1}^{n} x_{2i}\cdot x_{ni} = \sum_{i=1}^{n} y_i\cdot x_{2i} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0\sum_{i=1}^{n} x_{mi} + a_1\sum_{i=1}^{n} x_{1i}\cdot x_{mi} + a_2\sum_{i=1}^{n} x_{2i}\cdot x_{mi} \ldots + a_m\sum_{i=1}^{n} x_{mi}^2 = \sum_{i=1}^{n} y_i\cdot x_{mi} \end{cases} \tag{7.5}$$

где количество уравнений совпадает с числом переменных *(m+1)*, а суммирование осуществляется по выборке из *n* единиц.

Эта система может быть решена либо каким-нибудь из вычислительных методов линейной алгебры (например, методом Жордана-Гаусса), либо при помощи компьютера (в частности, такая функция реализуется в MS Excel).

Для множественной регрессионной модели актуален вопрос о том, какова сила влияния различных факторов на значение зависимой переменной. Для этого используются два основных метода. Первый основан на построении регрессионной модели в *стандартизированной форме,* второй – на расчете *частных коэффициентов эластичности*.

Для преобразования модели к стандартизованной форме осуществляют переход к новым переменным:

$$x'_{ij} = \frac{x_{ij} - \bar{x}}{s_{x_j}} \tag{7.6}$$

уравнение множественной регрессии примет вид:

$$y' = \beta_1 x'_1 + \beta_2 x'_2 + \ldots + \beta_m x'_m + \varepsilon = \sum_{j=1}^{m} \beta_j x'_j + \varepsilon \tag{7.7}$$

Оценив параметры этой модели по МНК, по значениям параметров β_j при каждой переменной мы можем сделать вывод о сравнительной силе влияния различенных факторов. Большее значение коэффициента говорит о большей силе влияния соответствующего фактора на зависимую переменную, положительное значение – о прямом влиянии, отрицательное – об обратном.

Частные коэффициенты эластичности рассчитываются по формулам:

$$Y_j' = a_j \frac{\overline{x}_j}{\overline{y}} \tag{7.8}$$

Коэффициент эластичности показывает, на сколько процентов изменяется зависимая переменная при изменении зависимой на 1 процент и неизменности действия прочих факторов.

Оценка качества построенной модели проводится по тем же этапам, что и для парной регрессии:

1. Анализ адекватности модели в целом
2. Анализ точности определения оценок коэффициентов регрессии (расчет их дисперсии и стандартного отклонения)
3. Проверка статистической значимости коэффициентов регрессии
4. Интервальная оценка коэффициентов регрессионного уравнения
5. Определение доверительных интервалов для зависимой переменной

Логика всех этапов аналогична парной регрессии.

Задания

1. Запишите спецификацию модели множественной линейной регрессии:
   a) рыночного спроса
   b) рыночного предложения
   c) совокупных расходов в кейнсианской модели «доходы - расходы»
   d) рациональных ожиданий
   e) трансакционных издержек
2. Проведите линеаризацию следующих моделей множественной регрессии:
   a) $y = \alpha_0 \alpha_1^{x_1} \alpha_2^{x_2} \alpha_3^{x_3} \varepsilon$
   b) $y = \alpha_0 x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} \varepsilon$
   c) $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 \sin(\mathrm{x}_2) + \alpha_3 \cos(\mathrm{x}_3) + \varepsilon$
3. По данным таблицы построить матрицу корреляции и разработать модель множественной регрессии, которая анализирует факторы, влияющие на объемы продаж в магазинах:

№	Qd (объем спроса)	P (цена товара, руб.)	L (расстояние от центра города, в км)	$P_з$. (цена товара-заменителя, руб.)	t (время существования магазина, лет)
1	25	2.5	1	10	18
2	30	2.3	7	9	20
3	35	2.4	25	10	20
4	35	2.2	10	9	12
5	37	2.4	16	8	22

6	42	2.1	8	9	8
7	50	2.0	25	8	30
8	70	2.0	10	10	12
9	72	1.8	20	8	20
10	74	1.9	15	7	30
11	80	2.1	10	9	8
12	110	1.8	8	8	14

Оцените качество полученной в предыдущем задании модели. Оцените статистическую значимость коэффициентов линейной регрессионной модели $Q_d = f(P, L, P_{в3/3}, t)$ из задания 3.

Литература: [1, 2, 6].

ЛАБОРАТОРНАЯ РАБОТА 8

Множественный регрессионный анализ

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: прогрессивные информационные технические средства и технологии.

Уметь: применять в работе математические методы и инструменты.

Владеть: способностью применять в работе математические методы и инструменты с целью решения стандартных и нестандартных профессиональных задач.

Актуальность темы: объясняется тем, что для множественной регрессионной модели актуален вопрос о том, какова сила влияния различных факторов на значение зависимой переменной. Для этого используются два основных метода. Первый основан на построении регрессионной модели в стандартизированной форме, второй - на расчете частных коэффициентов эластичности.

Теоретическая часть

1. Оценка адекватности модели в целом осуществляется на основе расчета *коэффициента детерминации* R^2 и *скорректированного коэффициента детерминации* $\hat{R}^2$, рассчитываемого по формуле:

$$\hat{R}^2 = 1 - \frac{\sum \frac{e_i^2}{n-m-1}}{\sum \frac{(y_i - \overline{y})^2}{n-1}} = 1 - \frac{n-1}{n-m-1} \cdot \frac{\sum e_i^2}{\sum (y_i - \overline{y})^2} \tag{8.1}$$

После несложных преобразований получаем выражение для $\hat{R}^2$ через R^2:

$$\hat{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-m-1} \tag{8.2}$$

из последней формулы видно, что скорректированный коэффициент детерминации меньше обычного коэффициента детерминации и, следовательно, является более строгим показателем связи, чем R^2.

Отметим, что корректировка может производиться, только если выполняется соотношение:

$$\frac{n-m}{m} \leq 20 \tag{8.3}$$

После определения значения коэффициента детерминации следует проанализировать его статистическую значимость. Статистическая значимость проверяется путем проверки гипотезы о равенстве коэффициента детерминации 0. Если гипотеза отвергается, то делается вывод о том, что коэффициент детерминации отличен от 0 и статистически значим. Для проверки используют *F-статистику*:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n-m-1}{m} \tag{8.4}$$

Полученное фактическое значение сравнивают с критическим $F_{\alpha;\, m;\, n-m-1}$, если оно оказывается больше критического, то нулевая гипотеза отвергается и делается вывод о статистической значимости коэффициента детерминации и существенности построенной модели. В противном случае модель нельзя использовать на практике.

Для множественной регрессии оценка качества модели в целом также может осуществляться с использованием средней ошибки аппроксимации:

$$\overline{\varepsilon} = \frac{1}{n} \sum_{i=1}^{m} \frac{|y_i - \hat{y}_i|}{|y_i|} \tag{8.5}$$

Если значение $\overline{\varepsilon}$ превышает 0,15 (15%), то модель недостаточно хорошо описывает фактические данные.

2. Расчет дисперсии коэффициентов регрессионного уравнения. Точный расчет производится с использованием элементов матричной алгебры и отводится на самостоятельное изучение. Приближенное же вычисление дисперсии можно осуществить по формуле:

$$\sigma_{a_i}^2 = \frac{\sigma_y^2(1-\mathrm{R}^2)}{\sigma_{x_i}^2 \cdot n \cdot (1-\mathrm{R}_i^2)} \tag{8.6}$$

где R_i –коэффициент полной корреляции *i-й* переменной с остальными независимыми переменными (может быть определен как корень квадратный коэффициента детерминации регрессионной модели $x_i = f(\mathrm{x}_1, \ldots, x_j, \ldots \mathrm{x}_m) + \varepsilon$, где *j* не равно *i*).

Рассмотрим одну из существенных проблем, возникающих при применении множественного регрессионного анализа – *мультиколлинеарности*. Под мультиколлинеарностью понимают тесную линейную взаимосвязь объясняющих переменных (рис. 8.1 и 8.2). Термин мультиколлинеарность введен Р. Фришем.

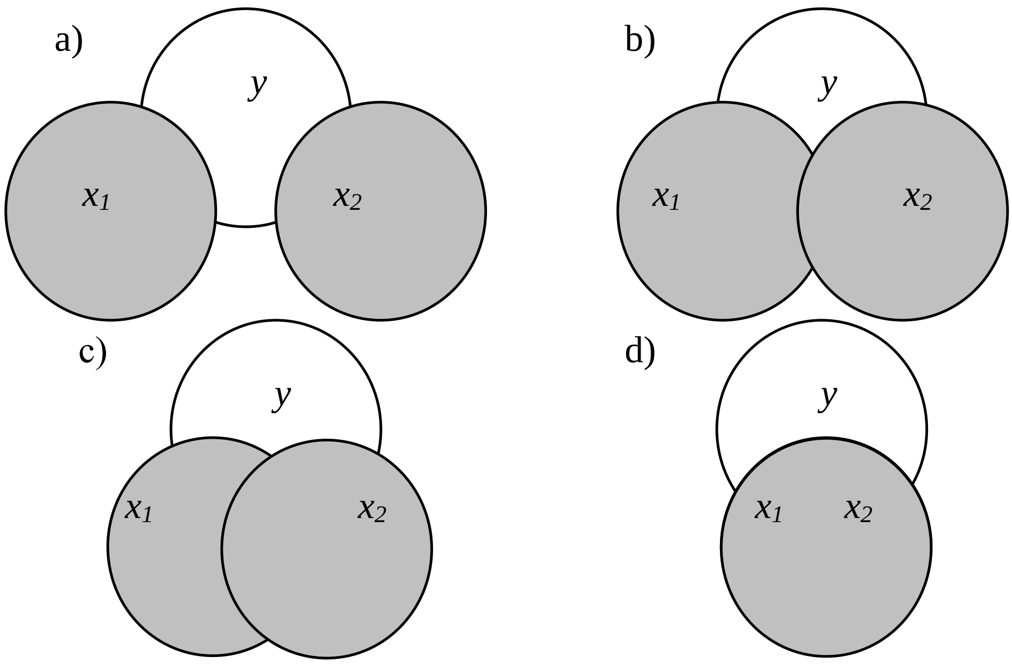


Рис. 8.1 Мультиколлинеарность между x_1 и x_2
a) мультиколлиенеарности нет; b) умеренная мультиколлиенарность;
c) сильная мультиколлинеарность;
d) совершенная мультиколлинеарность

В случаях, проиллюстрированных на рис. 8.1 a) и b) в модель можно включать обе независимые переменные, а в случаях c) и d) –только одну из двух переменных. Проблема мультиколлинеарности связана, прежде всего, со следующими ее *последствиями* (рис. 8.2).

Причины мультиколлинеарности:

«почти» функциональная связь между независимыми переменными вследствие:

постоянства суммы значений переменных

дублирующего экономического смысла переменных

того, что одна переменная является частью другой

неверная спецификация модели (включаются переменные, характеризующие одну сторону явления)

Последствия мультиколлинеарности:

высокие значения дисперсии оценок коэффициентов и ухудшение точности их интервальных оценок

чувствительность оценок коэффициентов к изменениям исходных данных

сложность оценки влияния каждой из объясняющих переменных на объясняемую

получение неверного знака у коэффициента перед объясняющей переменной

Методы обнаружения мультиколлинеарности:

анализ значений коэффициентов корреляции между объясняющими переменными

сопоставление коэффициента детерминации и статистической значимости коэффициентов в модели

анализ вспомогательной регрессии

анализ определителя матрицы корреляции независимых

Методы устранения мультиколлинеарности:

исключение коррелированных переменных из модели

проведение нового наблюдения

изменение спецификации модели

использование предварительной информации о значениях параметров

преобразования переменных

Рис. 8.2 Причины, последствия, методы обнаружения и устранения мультиколлинеарности

1. *Высокие значения дисперсии оценок коэффициентов*, что приводит к ухудшению точности их интервальных оценок, а также уменьшению t-статистик коэффициентов, что может привести к неправильному выводу о несущественности влияния независимой переменной на зависимую.

2. Оценки коэффициентов, полученные по МНК, становятся очень *чувствительными к изменениям исходных данных*.

3. Затрудняется *измерение вклада каждой из объясняющих переменных* в объясняемую уравнением регрессии дисперсию зависимой переменной.

4. Возможно *получение неверного знака у коэффициента* перед объясняющей переменной.

В то же время при достаточно высоких значениях коэффициента детерминации в моделях, построенных для цели прогнозирования, мультиколлинеарность не является достаточно серьезной проблемой, и построенные модели вполне могут быть использованы на практике.

Проблема мультиколлинеарности обязательно должна быть решена в том случае, если целью модели является анализ характера влияния различных факторов на зависимую переменную.

Мультиколлиеарность можно определить при помощи следующих *методов*:

1. *Анализ значений коэффициентов корреляции между объясняющими переменными.* Высокие коэффициенты корреляции между объясняющими переменными

2. *Сопоставление коэффициента детерминации и статистической значимости коэффициентов в модели*. Коэффициент детерминации модели достаточно высок, но некоторые из коэффициентов в модели статистически незначимы.

3. *Анализ вспомогательной регрессии* – регрессии между объясняющими переменными. Для обнаружения регрессионной зависимости между объясняющими переменными строятся регрессионные модели типа:

$$x_i = f(\mathrm{x}_1, \ldots, x_j, \ldots \mathrm{x}_m) + \varepsilon, \qquad (8.7)$$

где j не равно i

для каждой объясняющей переменной.

Затем при помощи критерия Фишера (F-статистики) проверяется ее статистическая значимость:

$$F_i = \frac{R_i^2}{1-R_i^2} \cdot \frac{n-m}{m-1} F_i = \frac{R_i^2}{1-R_i^2} \cdot \frac{n-m}{m-1} \qquad (8.8)$$

Полученное значение сравнивается с критическим $F_{\alpha;\, m-1;\, n-m}$. Если значение F_i оказывается больше критического, то делается вывод о том, что i-я независимая переменная является линейной комбинацией других и, следовательно, в модели присутствует мультиколлинеарность.

4. *Анализ определителя матрицы корреляции независимых переменных*.

Этот метод основан на том, что матрица, составленная из коэффициентов корреляции между объясняющими переменными, в случае отсутствия мультиколлинеарности имела бы определитель, равный единице (8.9):

$$\det M = \begin{vmatrix} r_{x_1x_1} & r_{x_1x_2} \dots & r_{x_1x_m} \\ r_{x_2x_1} & r_{x_2x_2} \dots & r_{x_2x_m} \\ \dots\dots & \dots\dots & \dots\dots \\ r_{x_mx_1} & r_{x_mx_2} \dots & r_{x_mx_m} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{vmatrix} = 1 \tag{8.9}$$

При существовании мультиколлинеарности коэффициенты отличны от 0 (изменяются в пределах от –1 до 1), и определитель матрицы становится меньше. В случае совершенной мультиколлинеарности он равен 0:

$$\det M = \begin{vmatrix} r_{x_1x_1} & r_{x_1x_2} \dots & r_{x_1x_m} \\ r_{x_2x_1} & r_{x_2x_2} \dots & r_{x_2x_m} \\ \dots\dots & \dots\dots & \dots\dots \\ r_{x_mx_1} & r_{x_mx_2} \dots & r_{x_mx_m} \end{vmatrix} = \begin{vmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{vmatrix} = 0 \tag{8.10}$$

Обнаружить мультиколлинеарность можно, проверив статистическую гипотезу по поводу равенства 1 определителя этой матрицы:

H_0: det M = 1 (8.11)

Проверка осуществляется на основе критерия χ^2(хи –квадрат). Доказано, что величина:

$$\chi^2 = n - 1 - \frac{1}{6}(2\,\mathrm{m} + 5)\,\mathrm{lg}\,\mathrm{det}\,\mathrm{M} \tag{8.12}$$

имеет распределение χ^2 с $\left(\frac{n \cdot (\mathrm{n}-1)}{2}\right)$ степенями свободы. Расчетное значение сравнивается с табличным, и если расчетное значение оказывается больше, то считается, что мультиколлинеарность имеет место.

Рассмотрим *методы устранения* мультиколлинеакрности.

1. *Исключение коррелированных переменных из модели.* Это наиболее простой способ борьбы с мультиколлинеарностью. Однако в этом случае возможны серьезные проблемы, связанные с тем, что полученные по упрощенной модели оценки будут смещенными.

2. *Получение дополнительных данных или проведение нового наблюдения.* Часто мультиколлинеарность проявляется вследствие неполноты данных, и при расширении выборки существенно уменьшается. Однако этот подход связан со значительными издержками, и, кроме того, может быть связан с появлением такого нежелательного явления, как автокорреляция.

3. *Изменение спецификации модели.* Может быть осуществлено как при помощи изменения аналитического выражения модели, так и путем добавления новых переменных, оказывающих существенное влияние на зависимую переменную. Этот метод целесообразно применять, если добавляемая переменная является «полезной», то есть существенно улучшает качество модели.

4. *Использование предварительной информации о значениях некоторых параметров*. Иногда значения некоторых неизвестных параметров модели могут быть определены по пробным выборочным наблюдениям, тогда мультиколлинеарность может быть устранена путем установления значений параметра у одной коррелирующих переменных. Ограниченность метода – в сложности получения предварительных значений параметров с высокой точностью.

5. *Преобразование переменных.* Для устранения мультиколлинеарности можно преобразовать переменные, например, путем линеаризации или получения относительных показателей, а также перехода от номинальных к реальным показателям (особенно в макроэкономических исследованиях).

При построении модели множественной регрессии с точки зрения обеспечения ее высокого качества возникают следующие вопросы:

1. Каковы признаки качественной модели?
2. Какие ошибки спецификации могут быть?
3. Каковы последствия ошибок спецификации?
4. Какие существуют методы обнаружения и устранения ошибок спецификации?

Рассмотрим основные *признаки качественной модели* множественной регрессии:

1. *Простота.* Из двух моделей примерно одинаковых статистических свойств более качественной является та, которая содержит меньше переменных, или же более простая по аналитической форме.

2. *Однозначность.* Метод вычисления коэффициентов должен быть одинаков для любых наборов данных.

3. *Максимальное соответствие.* Этот признак говорит о том, что основным критерием качества модели является коэффициент детерминации, отражающий объясненную моделью вариацию зависимой переменной. Для практического использования выбирают модель, для которой расчетное значение F-критерия для коэффициента детерминации б четыре раза больше табличного.

4. *Согласованность с теорией.* Получаемые значения коэффициентов должны быть интерпретируемы с точки зрения экономических явлений и процессов. К примеру, если строится линейная регрессионная модель спроса на товар, то соответствующий коэффициент при цене товара должен быть отрицательным.

5. *Хорошие прогнозные качества.* Обязательным условием построения качественной модели является возможность ее использования для прогнозирования.

Одной из основных ошибок, допускаемых при построении регрессионной модели, является *ошибка спецификации (рис. 8.3)*. Под ошибкой спецификации понимается неправильный выбор функциональной формы модели или набора объясняющих переменных.

Различают следующие *виды ошибок спецификации*:

1. Невключение в модель полезной (значимой) переменной.
2. Добавление в модель лишней (незначимой) переменной

3. Выбор неправильной функциональной формы модели

Последствия ошибки *первого вида (невключение в модель значимой переменной)* заключаются в том, что полученные по МНК оценки параметров являются смещенными и несостоятельными, а значение коэффициента детерминации значительно снижаются.

Рис. 8.3 Ошибки спецификации и свойства качественной регрессионной модели

При *добавлении в модель лишней переменной (ошибка второго вида)* ухудшаются статистические свойства оценок коэффициентов, возрастают их дисперсии, что ухудшает прогнозные качества модели и затрудняет содержательную интерпретацию параметров, однако по сравнению с другими ошибками ее последствия менее серьезны.

Если же осуществлен *неверный выбор функциональной формы модели*, то есть допущена ошибка третьего вида, то получаемые оценки будут смещенными, качество модели в целом и отдельных коэффициентов будет невысоким. Это может существенно сказаться на прогнозных качествах модели.

Ошибки спецификации первого вида можно обнаружить только по невысокому качеству модели, низким значениям R^2.

Обнаружение ошибок спецификации второго вида, если лишней является только одна переменная, осуществляется на основе расчета *t - статистики* для коэффициентов. При лишней переменной коэффициент будет статистически незначим.

Если же таких переменных несколько, целесообразно прибегнуть к сравнению значений коэффициентов детерминации модели до и после исключения из модели переменных, которые считаются лишними, при помощи расчета *F-критерия* по формуле:

$$F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - m_1 - 1}{m_1 - m_2} \tag{8.13}$$

где m_1 – число объясняющих переменных в первоначальном уравнении, m_2 – число объясняющих переменных в уравнении после отброса лишних переменных.

Полученное значение сравнивается с критическим $F_{\alpha; m1-m2; n-m1-1}$. Если расчетное значение меньше, то считается, что исключенные из модели переменные являются лишними.

Ошибки третьего вида можно обнаружить только при помощи содержательной интерпретации модели или визуально анализируя данные или по наличию гетероскедастичности (см. тему 7).

Комплексный анализ ошибок спецификации можно провести, выполнив один или несколько из следующих тестов:

1) Тест Рамсея (Regression specification error test – RESET);
2) Тест максимального правдоподобия (The Likelihood Ratio test);
3) Тест Валда (The Wald test);
4) Тест множителя Лагранжа (The Lagrange multiplier test);
5) Тест Хаусманна (The Hausmann test)
6) Преобразование Бокса-Кокса (Box-Cox transformation)

Задания

1. Определите наличие (отсутствие) мультиколлинеарности в модели из задания 3 различными методами.
2. Оцените параметры ПФ Кобба-Дугласа по следующим данным:

№ п/п	Q	L	K
1.	1120	80	30
2.	1000	75	30

3.	1400	70	45
4.	1520	80	43
5.	1550	90	42
6.	1600	82	45
7.	1380	75	38
8.	1500	80	41
9.	1400	77	40
10.	1280	90	20

3. Оцените параметры линейной производственной функции, производственной функции Кобба-Дугласа и производственной функции Леонтьева по следующим эмпирическим данным. Какая функция точнее описывает фактическую зависимость?

Объемы производства (Q), млн. руб.

Объем используемого труда (L), человек / Объем используемого капитала (K), млн. руб.	400	420	440	460	480	500	520
10	78,8	79,5	79,2	81,5	82,7	81,7	84,3
12	91,1	92,1	92,0	92,7	93,8	97,4	98,0
14	101,7	104,1	104,8	104,7	106,1	110,3	108,4
16	112,8	115,3	115,2	116,8	120,1	121,6	120,4
18	122,9	125,7	128,1	129,7	131,4	133,5	135,8
20	134,5	135,7	139,0	141,9	142,1	142,1	146,5
22	144,3	147,6	147,3	153,0	154,7	155,5	156,3

4. Если в результате построенной вспомогательной регрессии в модели (m=5, n=100) получились следующие результаты:
   $R_1 = 0,77$
   $R_2 = 0,28$
   $R_3 = 0,56$
   $R_4 = 0,92$
   $R_5 = 0,07$
   Какие переменные следует исключить из модели и почему?
   Литература: [1, 2, 5].

ЛАБОРАТОРНАЯ РАБОТА 9

Регрессионные модели с переменной структурой

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: культуру мышления и - способы обобщения информации.

Уметь: анализировать социально значимые проблемы и процессы, происходящие в обществе, и прогнозировать возможное их развитие в будущем и уметь строить стандартные теоретические и эконометрические модели.

Владеть: культурой мышления и современными методами сбора, обработки и анализа информации.

Актуальность темы: определяется тем, что часто возникает необходимость учета в регрессионной модели факторов, не имеющих количественного выражения. Например, при анализе динамики производства по месяцам следует учитывать сезонную компоненту.

Теоретическая часть

Часто возникает необходимость учета в регрессионной модели факторов, не имеющих количественного выражения. Например, при анализе динамики производства по месяцам следует учитывать сезонную компоненту. При анализе потребительских предпочтений может возникнуть задача учета пола потребителя, его отношения к определенной социальной группе. При регрессионном анализе предприятий может возникнуть потребность учета его формы собственности, отрасли, к которой предприятие относится. Модели, в которых учитываются подобные атрибутивные признаки, называются *моделями с переменной структурой*.

Для решения проблемы учета таких факторов можно:

1. Составить несколько регрессионных моделей, каждая из которых отражает зависимость в качественно однородной среде. Это приведет к значительному усложнению как непосредственно расчетов, так и интерпретации результатов.

2. Добавить в модель *фиктивные переменные*. Эти переменные могут принимать только два значения: 1, если наблюдается определенное значение признака, и 0, если оно не наблюдается. При этом количество переменных для включения одного признака будет на 1 меньше числа значений, принимаемых этим признаком.

Пример:

Рассмотрим регрессионную модель влияния на рентабельность активов различных факторов. Если необходимо учесть влияние фактора «организационно-правовая форма», то следует добавить следующие фиктивные переменные:

$$z_1 = \begin{cases} 1, & \text{если предприятие} - \text{ОАО} \\ 0, & \text{в противном случае} \end{cases} \tag{9.1}$$

$$z_2 = \begin{cases} 1, & \text{если предприятие} - \text{ЗАО} \\ 0, & \text{в противном случае} \end{cases} \tag{9.2}$$

$$z_3 = \begin{cases} 1, & \text{если предприятие} - \text{ООО} \\ 0, & \text{в противном случае} \end{cases} \tag{9.3}$$

(предполагается, что круг рассматриваемых предприятий ограничивается только тремя организационно-правовыми формами: ОАО, ЗАО и ООО).

Как несложно заметить, число фиктивных переменных можно сократить, так как всегда будет выполняться равенство:

$$z_3 = 1 - z_1 - z_2 \tag{9.4}$$

В общем случае построенная с включением фиктивных переменных регрессионная модель может иметь вид:

а) $y = f(z_1, z_2 \ldots z_k) + \varepsilon$ – модели, содержащие только фиктивные объясняющие переменные, представляющие собой кусочно-постоянные функции (они называются *ANOVA* – моделями или моделями дисперсионного анализа)

б) $y = f(x_1, x_2 \ldots x_m, z_1, z_2 \ldots z_k) + \varepsilon$ – модели, содержащие и количественные, и фиктивные объясняющие переменные, и называемые *ANCOVA* – моделями или моделями ковариационного анализа). Такие модели имеют вид множества функций, непрерывных на своей области определения. На практике чаще рассматриваются модели *ANCOVA*, в которых фиктивные переменные играют уточняющую роль.

Можно выделить три основных направления использования моделей с переменной структурой: моделирование *сезонных колебаний,* моделирование *институциональных изменений и* учет влияния на результативный признак *качественных факторов* (рис. 9.1).

Учет сезонных колебаний осуществляется путем добавления 3 (если имеются квартальные данные) или 11 (если информация представлена по месяцам) фиктивных переменных. Учет институциональных изменений осуществляется путем добавления переменной, характеризующей данные как «до» и «после» изменений. Например, если рассматривается хозяйственная деятельность предприятия, и необходимо учесть влияние смены директора на его эффективность, то в модель необходимо добавить переменную z, такую, что:

$$z = \begin{cases} 1, & \text{если данные относятся к периоду до смены директора} \\ 0, & \text{в противном случае} \end{cases}$$

При применении модели в анализе качественных признаков (таких, как пол, социальный статус, форма собственности и т.п.) использование фиктивных переменных осуществляется по стандартной схеме.

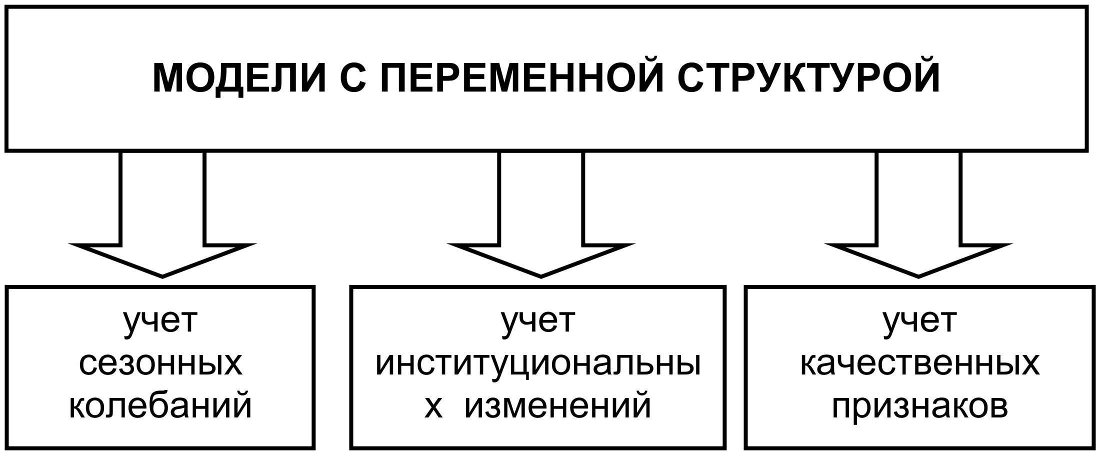


Рис. 9.1 Использование моделей с переменной структурой

Спецификация модели с переменной структурой обычно, в части фиктивных переменных, имеет вид линейной функции:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_{m1} x_{m1} + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{m2} z_{m2} + \varepsilon \qquad (9.5)$$

Однако в ряде случаев возможно использование альтернативной фиктивной переменной (т.е. принимающей только два значения) в мультипликативной функции вида:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_{m1} x_{m1} + z \cdot (\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{m1} x_{m1}) + \varepsilon \qquad (9.6)$$

Спецификация (9.6) используется, когда явно можно предположить, что изменение фиктивной переменной отразится не только на изменении начального уровня явления и параллельном сдвиге линии регрессии, но и на ее наклоне. Такое предположение возможно, в частности, когда рассматривается зависимость заработной платы от стажа и пола сотрудников, или же при учете институциональных изменений.

Последовательность решения задач эконометрического анализа моделей с переменной структурой после проведения спецификации модели не отличается от последовательности, применяемой в решении обычных задач множественной регрессии: это определение оценок неизвестных параметров при помощи МНК и оценка качества модели.

Обратите внимание, что для применения МНК необходимо выполнение его предпосылок, изложенных в теме 4.

Для определения оценок параметров в модели (9.5) можно сразу использовать МНК. Если же используется модель вида (9.6), то следует либо рассматривать несколько регрессий, в которых значения фиктивных переменных заданы, либо преобразовать переменные путем замены (в случае одной фиктивной переменной):

$$\begin{cases} z \cdot x_1 = x_1' \\ z \cdot x_2 = x_2' \\ z \cdot x_{m1} = x_{m1}' \end{cases} \qquad (9.7)$$

Тогда спецификация модели изменится, и она примет вид обычной множественной регрессии:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_{m1} x_{m1} + \beta_1 x_1' + \beta_2 x_2' + \ldots + \beta_{m1} x_{m1}' + \varepsilon \qquad (9.8)$$

Определение параметров в такой модели осуществляется при помощи обычного МНК, затем, в случае необходимости, можно вернуться к исходной форме.

Тесты

1. Для учета сезонных колебаний по квартальным данным число фиктивных переменных должно быть равно:
   a) 1
   b) 3
   c) 4
   d) 5
2. Модели, в которых независимые переменные могут быть как количественные, так и фиктивные, называются:
   a) ANOVA
   b) ANCOVA
   c) LPM
   d) LOGIT
3. Какие из моделей с фиктивной зависимой переменной допускают ситуацию, когда значение фиктивной переменной выйдет за границы [0;1]:
   a) ANOVA
   b) ANCOVA
   c) LPM
   d) PROBIT
4. Сумма значений m фиктивных переменных, характеризующих один качественный признак, равна:
   a) 0
   b) 1
   c) m – 1
   d) m
5. Мультипликативная форма модели с переменной структурой будет иметь вид:
   a) $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \varepsilon$
   b) $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + z \cdot (\beta_1 x_1 + \beta_2 x_2) + \varepsilon$
   c) $y = \alpha_0 \cdot \alpha_1 x_1 \cdot \alpha_2 x_2 + z + \varepsilon$
   d) $y = \alpha_0 \cdot \alpha_1 x_1 \cdot \alpha_2 x_2 \cdot z + \varepsilon$

Литература: [1, 2, 5].

ЛАБОРАТОРНАЯ РАБОТА 10

Регрессионные модели с переменной структурой

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: культуру мышления и - способы обобщения информации.

Уметь: анализировать социально значимые проблемы и процессы, происходящие в обществе, и прогнозировать возможное их развитие в будущем и уметь строить стандартные теоретические и эконометрические модели.

Владеть: культурой мышления и современными методами сбора, обработки и анализа информации.

Актуальность темы: определяется тем, что часто возникает необходимость учета в регрессионной модели факторов, не имеющих количественного выражения. Например, при анализе динамики производства по месяцам следует учитывать сезонную компоненту.

Теоретическая часть

Оценка качества моделей с переменной структурой осуществляется по тому же алгоритму, что и для обычной множественной регрессии.

Общую характеристику качеству построенной модели может дать *коэффициент детерминации* R^2 (поскольку модель множественной регрессии, то рекомендуется использовать *скорректированный коэффициент детерминации* $\hat{R}^2$, рассчитываемый по формуле (8.7)). Статистическая значимость коэффициента детерминации может быть оценена при помощи F-критерия.

В том случае, если использовалась мультипликативная форма модели (9.5) с одной фиктивной переменной, для оценки качества модели может быть использован *тест Чоу*.

Тест Чоу основан на F-критерии. Для его проведения необходимо построить регрессионную модель без участия фиктивной переменной (общую модель), рассчитать сумму квадратов отклонений фактических значений от общей модели. F-критерий рассчитывается как соотношение величины, на которую уменьшилась дисперсия в модели с переменной структурой по сравнению с общей моделью и дисперсии модели переменной структурой:

$$F = \frac{\sum e^2 - \sum e_z^2}{\sum e_z^2} \cdot \frac{n-2m-1}{m+1} \tag{10.1}$$

где $\sum e^2$ – сумма квадратов случайных отклонений в общей модели

$\sum e_z^2$ – сумма квадратов случайных отклонений в регрессионной модели с переменной структурой

Полученное фактическое значение сравнивается с критическим $F_{\alpha;\, m+1;\, n-2m-1}$. Если расчетное значение оказывается больше фактического, то считается, что использование фиктивной переменной целесообразно.

Также фиктивная зависимая переменная может быть использована для интерпретации влияния на качественный признак количественных факторов. В этом случае говорят о *фиктивной зависимой переменной.* Различают несколько видов таких моделей:

1. *Линейная вероятностная модель* (linear probability model – *LPM*)
2. *Модели бинарного выбора* (*LOGIT* и *PROBIT* модели).

Спецификация линейной вероятностной модели имеет вид (10.1) – если все объясняющие переменные описывают количественные факторы, или (8.5) – если в модели присутствуют фиктивные объясняющие переменные. Однако в этом случае зависимая переменная *y* может принимать только одно из двух значений:

$$y = \begin{cases} 1, & \text{если признак наблюдается} \\ 0, & \text{в противном случае} \end{cases} \tag{10.2}$$

После расчета параметров модели по МНК расчетные значения зависимой переменной интерпретируются как вероятность того, что признак, характеризуемый *y*, будет наблюдаться. Однако применение этого метода ограничено тем, что:

1) не выполняется ряд предпосылок МНК (случайные остатки не распределены по нормальному закону, дисперсия остатков непостоянна);

2) возможна ситуация, когда расчетное значение у будет находиться вне области допустимых значений: [0,1];

3) с содержательной точки зрения линейное изменение вероятности при изменении влияния факторов некорректно.

Для преодоления последних двух условий разработаны и используются модели бинарного выбора, среди которых выделяют *LOGIT* и *PROBIT* модели.

Рассмотрим суть *LOGIT* модели. В ней в качестве функции, характеризующей вероятность того, что *y* примет значение 1, рассматривается логистическая функция. Эта функция принимает значения строго от 0 до 1:

$$p_i = P(y = 1x_i) = \frac{1}{1+e^{-z_i}} \tag{10.3}$$

где $z_i = \alpha_0 + \alpha_1 x_i$

Однако для применения МНК необходимо выполнить преобразования, поскольку функция (5.11) не линейна относительно параметров. Из (10.3) следует:

$$1-p_i=1-\frac{1}{1+e^{-z_i}}=\frac{e^{-z_i}}{1+e^{-z_i}}=\frac{1}{1+e^{-z_i}} \qquad (10.4)$$

$$\frac{p_i}{1-p_i}=\frac{1+e^{z_i}}{1+e^{-z_i}}=e^{z_i} \qquad (10.5)$$

$$\ln\frac{p_i}{1-p_i}=z_i=\alpha_0+\alpha_1 x_i \qquad (10.6)$$

Последнее выражение линейно относительно параметров, однако к нему нельзя применить МНК. Причина этого в том, что для фактических данных неизвестными остаются значения p_i . Для их нахождения в случае сгруппированных данных возможно использование в качестве оценки относительную частость проявления $\tilde{p}_i=\frac{n_i}{n}$, в случае несгруппированных следует применять метод максимального правдоподобия.

Задания

1. Оцените качество модели с переменной структурой, если:

m = 4

n = 50

S_e^2 = 1200

S_y^2 = 10800

$S_{e(z)}^2$ = 1000

Есть ли необходимость в использовании фиктивных переменных?

2. По данным таблицы постройте LPM и LOGIT – модели и рассчитайте их параметры. Какова вероятность банкротства для предприятия если $К_{ал}$ = 0,002, $К_{тл}$ = 1,2, рентабельность в прошлом году – (- 0,12), в текущем – (+0,07). Насколько качественна построенная модель.

№ п/п	Группа по фин. устойчивости	$К_{ал}$	$К_{тл}$	R_{-1} (рентабельность в предшествующем периоде)	R (рентабельность в текущем периоде)	Кол-во предприятий	Кол-во банкротств
	Очень неустойчиво	0,001	0,92	-0,15	-0,12	12	7
	Достаточно неустойчиво	0,05	1,2	-0,25	-0,11	21	9
	Неустойчиво	0,001	2,3	-0,10	-0,08	20	8

	Недостаточно устойчиво	0,03	2,2	0,11	0,02	17	4
	Средней устойчивости	0,05	2,5	-0,01	0,05	21	5
	Достаточно устойчиво	0,08	2,7	0,10	0,12	25	2
	Устойчиво	0,10	2,8	0,25	0,22	31	2
	Чрезвычайно устойчиво	0,15	3,2	0,10	0,09	24	1

3. Какова вероятность того, что удастся сдать экзамен на отлично, потратив на подготовку 2 дня, если в соответствующей LOGIT-модели $z=-5+3x$, где x – количество дней, затраченных на подготовку. Сколько дней нужно потратить, чтобы на 90% быть уверенным в отличной оценке.

4. В городе N была проведено исследование «честный кондуктор». В течение недели 3 респондента «случайно» давали за проезд кондуктору больше мелких монет, чем составляла стоимость проезда. Цель исследования – определить силу и характер влияния на то, вернет ли кондуктор «лишние» деньги, различных факторов. В результате получены следующие результаты:

№п/п	Вернул ли кондуктор лишние деньги	Заполненность троллейбуса (автобуса), %	возраст кондуктора, лет	Респондент
1.	вернул	32,1	33	1
2.	не вернул	94,5	55	1
3.	вернул	61,1	55	1
4.	вернул	26,3	35	1
5.	не вернул	49,2	32	2
6.	не вернул	73,3	50	2
7.	не вернул	77,4	22	2
8.	вернул	22,1	25	2
9.	вернул	12,7	35	3
10.	вернул	38,2	29	3
11.	вернул	20,8	30	3
12.	не вернул	56,6	30	3

Литература: [1, 2, 5].

ЛАБОРАТОРНАЯ РАБОТА 11

Специфика построения динамических регрессионных моделей

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: культуру мышления и - способы обобщения информации.

Уметь: анализировать социально значимые проблемы и процессы, происходящие в обществе, и прогнозировать возможное их развитие в будущем и уметь строить стандартные теоретические и эконометрические модели.

Владеть: культурой мышления и современными методами сбора, обработки и анализа информации.

Актуальность темы: определяется тем, что большинство эконометрических моделей разработаны для анализа стационарного ряда, поскольку методы его анализа значительно проще, а нестационарный ряд достаточно легко приводится к стационарному путем исключения основной временной тенденции (тренда). Для нас особый интерес представляют факторные динамические модели, в которых в качестве объясняющих переменных используются переменные, принадлежащие различным периодам времени.

Теоретическая часть

Динамической регрессионной моделью называется регрессионная модель, в которой в качестве объясняющих переменных используются не только текущие, но и предшествующие значения, а также временной фактор. Динамические регрессионные модели строятся по данным из *временных рядов (рядов динамики)* – последовательности данных, характеризующих состояние объекта в различные моменты времени.

Различают стационарные и нестационарные временные ряды. В *стационарных временных рядах* вероятностные характеристики анализируемого показателя не меняются со временем. В таких рядах колебание зависимой переменной обусловлено действием всевозможных факторов, либо совершенно не связанных со временем (например, изменение величины экспорта во времени может зависеть от валютных курсов), либо взаимосвязь эта имеет периодический характер (величина экспорта может достаточно сильно колебаться в зависимости от времени года).

Другими словами, свойства строго стационарного временного ряда не меняются при изменении начала отсчета времени. В частности, из предположения о строгой стационарности временного ряда x_t следует, что закон распределения вероятностей случайной величины x_t не зависит от t, а значит, не зависят от t и все его основные числовые характеристики, в том числе: среднее значение $M(x_t) = \mu$ и дисперсия $D(x_t) = \sigma^2$.

В *нестационарных временных рядах* анализируемый признак имеет определяемую временем устойчивую тенденцию изменения. Соответственно, все характеристики нестационарного временного ряда зависят от фактора времени.

Динамические регрессионные модели бывают различных видов (см. рис. 11.1)

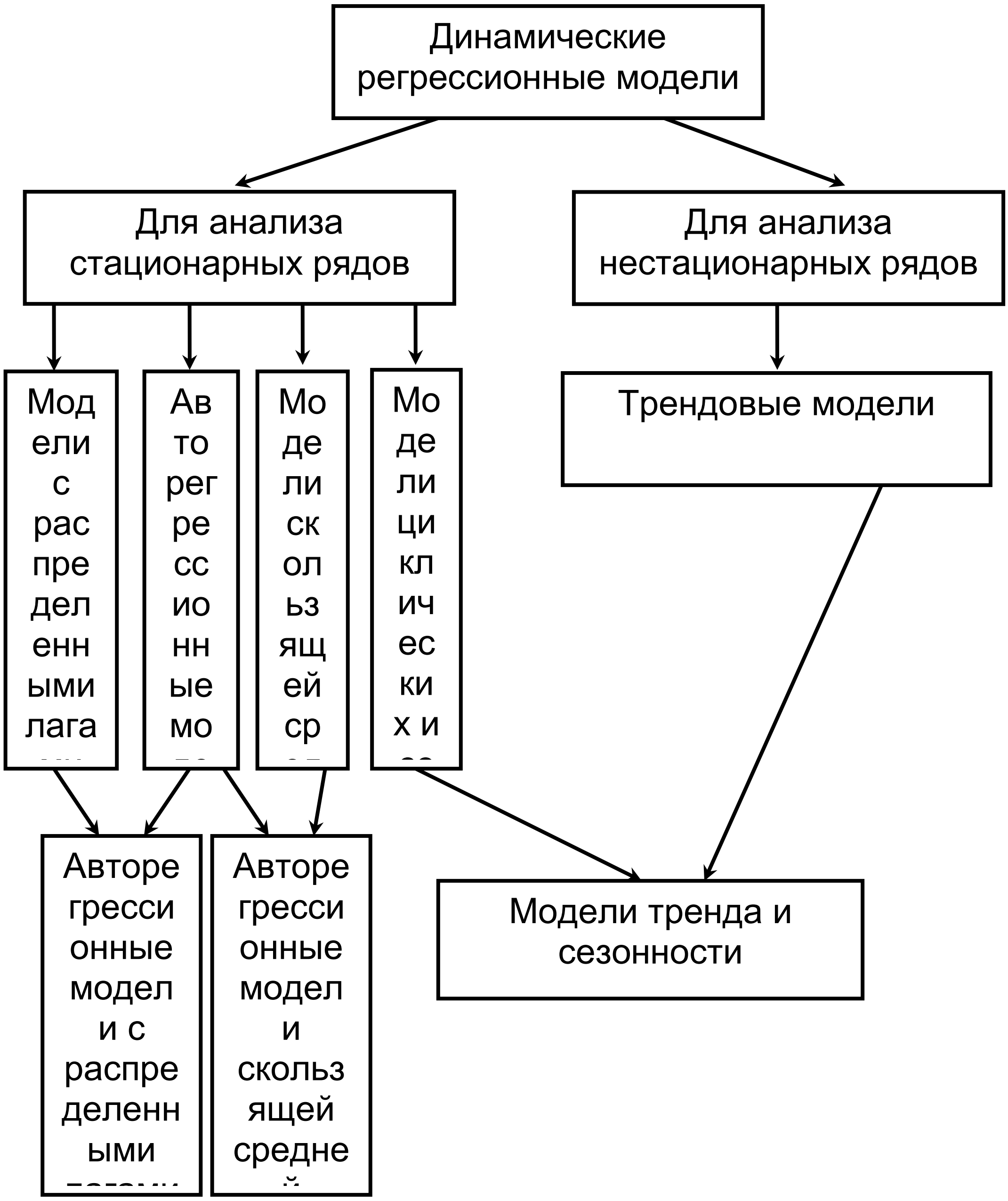


Рис. 11.1 Типология динамических регрессионных моделей

Большинство эконометрических моделей разработаны для анализа стационарного ряда, поскольку методы его анализа значительно проще, а нестационарный ряд достаточно легко приводится к стационарному путем исключения основной временной тенденции (*тренда*). Более подробно об анализе основной тенденции и колебательных (сезонных) процессов в ряде динамики можно прочитать в учебниках по статистике и в специализированной литературе.

Для нас особый интерес представляют *факторные динамические модели*, в которых в качестве объясняющих переменных используются переменные, принадлежащие различным периодам времени.

В том случае, если в качестве объясняющих переменных используются значения результативной переменной в предыдущие периоды, то такая модель называется *авторегрессионной*.

Переменные, отражающие значения факторов в предшествующие периоды, называются *лаговыми переменными*. Временные лаги (запаздывание реакции зависимой переменной на изменение факторов) существуют в экономике по следующим причинам:

1. *Психологические причины*. Сюда, в первую очередь, входит инерция в поведении людей, связанная с адаптивными ожиданиями.

2. *Технологические причины.* К примеру, действие фактора НТП не приведет к мгновенному росту, поскольку внедрение новых технологий – достаточно длительный по времени процесс. Точно также инвестиции, направленные в отрасль, дадут отдачу только через некоторое время.

3. *Институциональные причины.* Большинство соглашений в экономике имеют фиксированный срок действия, и отказаться от них, даже в случае изменения внешних факторов, затруднительно.

4. *Методы расчета анализируемых показателей.* Такие показатели, как уровень инфляции, уровень безработицы, денежный мультипликатор и т.п. изначально определяются как функции от переменных, принадлежащих к различным временным периодам

Для определения наличия лагов в авторегрессионных моделях используют *коэффициент автокорреляции*, который характеризует силу и направление связи между членами одного временного ряда, отстающими друг от друга на фиксированный временной период.

Статистической оценкой коэффициента автокорреляции является *выборочный коэффициент автокорреляции* (6.1):

$$r(\tau)=\frac{(n-\tau)\sum_{t=1}^{n-\tau} y_t y_{t+\tau}-\sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y_t^2-\left(\sum_{t=1}^{n-\tau} y_t\right)^2}\cdot\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y_{t+\tau}^2-\left(\sum_{t=1}^{n-\tau} y_{t+\tau}\right)^2}} \qquad (11.1)$$

где τ – временной лаг, для которого определяется корреляция.

Фактически выборочный коэффициент автокорреляции можно представить как коэффициент корреляции между членами одного временного ряда:

$$r(\tau)=r_{y_t y_{t-\tau}} \qquad (11.2)$$

Выборочный коэффициент автокорреляции обычно вычисляется для временных лагов, не превышающих четверти выборки, то есть $\tau \leq n/4$. График зависимости $r(\tau)$ называется *коррелограммой*. В модель обычно включают переменные, соответствующие лагам с высокими по абсолютному значению выборочными коэффициентами автокорреляции (рис. 11.2).

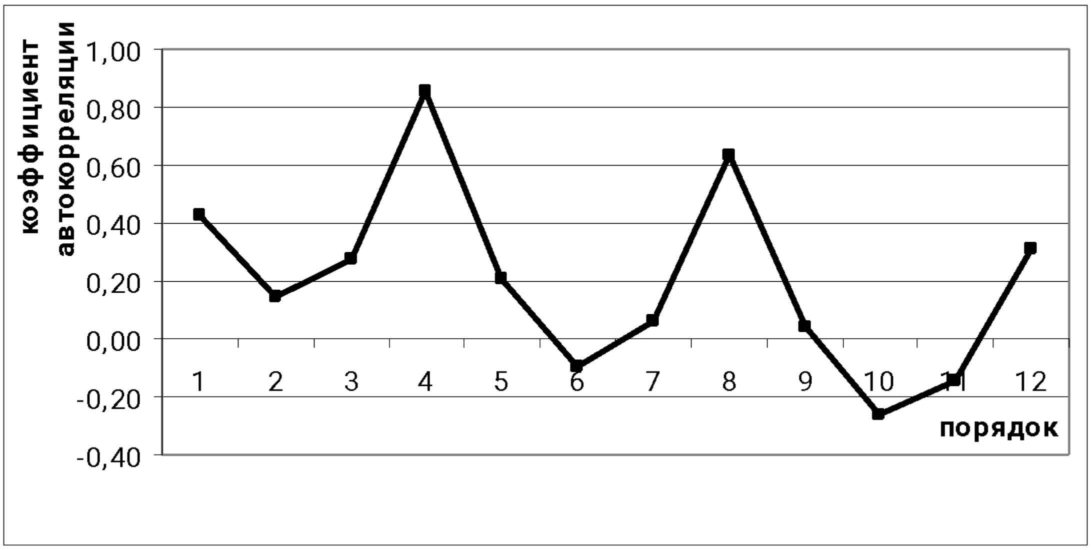


Рис. 11.2 Кореллограмма, построенная по квартальным данным (заметна сильная положительная корреляция между каждым четвертым кварталом)

С точки зрения спецификации факторные динамические модели, используемые для анализа стационарных рядов, бывают различных видов:

1. *Модели с распределенными лагами p-го порядка (DL(p))* – модели, в которых учитываются только лаги в независимых переменных:

$$y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} \ldots + \beta_p x_{t-p} + \varepsilon_t \tag{11.3}$$

2. *Авторегрессионная модель p-го порядка (AR(p))* – это динамическая регрессионная модель, в которой в качестве объясняющих переменных используются значения зависимой переменной за *p* предшествующих периодов.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} \ldots + \beta_p y_{t-p} + \varepsilon_t \tag{11.4}$$

3. *Модель скользящей средней q-го порядка (MA(q))* – это модель, в которой в качестве объясняющих переменных используются случайные ошибки в предыдущих временных периодах.

$$y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} \ldots + \gamma_q \varepsilon_{t-q} \tag{11.5}$$

4. *Авторегрессионная модель скользящей средней порядков p и q (ARMA(p,q))* – это модель, являющаяся объединением двух предыдущих, т.е. в ней в качестве объясняющих переменных используются как предыдущие значения ряда, так и предыдущие значения ошибок.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} \ldots + \beta_p y_{t-p} + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} \ldots + \gamma_q \varepsilon_{t-q} + \varepsilon_t \tag{11.6}$$

5. *Авторегрессионная модель с распределенными лагами порядка p и q (ADL(p,q))* – модели, в которых учитываются лаги и в независимых, и в зависимых переменных:

$$y_t = \beta_0 + \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \beta_2 y_{t-2} \ldots + \beta_p y_{t-p} + \varepsilon_t \tag{6.7}$$

Рассмотрим модели с лагами в независимых переменных (*DL(p)*). В том случае, если количество лагов конечно (например, для построения модели используются данные бухгалтерского баланса за 5 лет) модель может быть оценена при помощи обычного МНК для случая p независимых переменных.

Если же теоретически лагов может быть сколь угодно много (например, когда рассматривается фондовый рынок, на котором временные ряды для курсов акций могут быть построены с интервалом в 1 минуту), то для оценки неизвестных коэффициентов можно использовать два основных метода:

1. Метод последовательного увеличения количества лагов. Его применение осуществляется следующим образом.

1) Рассматривается модель с одним лагом: $y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$. Для нее вычисляются оценки и ошибки всех коэффициентов.

2) Количество лагов увеличивается на единицу $y_t = \alpha_0 + \alpha_1 x_{t-1} + \ldots + \alpha_p x_{t-p} + \varepsilon_t$ (до p) и оцениваются параметры вновь построенной модели.

3) Если коэффициент *(p–1)-й* лаговой переменной изменил знак, то последняя лаговая переменная отбрасывается, и делается вывод о том, что следует рассматривать модель с p лагами. Иначе продолжаем.

4) Если коэффициент *p-й* лаговой переменной статистически незначим, то делается вывод о том, что следует рассматривать модель с p лагами. Иначе продолжаем.

5) Возврат на шаг 2.

Основной проблемой применения этого метода является постоянное сокращение числа степеней свободы, из-за чего увеличиваются ошибки коэффициентов, а также часто возникающая проблема мультиколлинеарности.

2. Метод геометрической прогрессии

В этой модели предполагается, что коэффициенты при лагах в зависимой переменной убывают в геометрической прогрессии:

$\beta_t = \beta_0 \lambda^k$, где λ – коэффициент, характеризующий скорость убывания коэффициентов с увеличением лага ($0<\lambda<1$)

В этом случае спецификация модели будет иметь вид:

$$y_t = \alpha_0 + \beta_0 \lambda \cdot x_t + \beta_0 \lambda^2 \cdot x_{t-1} \ldots + \beta_p \lambda^p \cdot x_{t-p} + \varepsilon_t \tag{11.8}$$

то есть в модели будет 3 неизвестных параметра ($\alpha_0, \beta_0, \alpha_0, \beta_0, \lambda$).

Для расчета их значений используют два основных метода:

I. Произведя замену $z_t = \lambda \cdot x_t + \lambda^2 \cdot x_{t-1} \ldots + \lambda^p \cdot x_{t-p}$ мы получаем линейную парную регрессию:

$$y_t = \alpha_0 + \beta_0 z_t + \varepsilon_t \tag{11.9}$$

Как видно, значения z_t зависят от неизвестного параметра λ, поэтому, прежде чем оценить параметры регрессии (11.9), следует определить, какое значение λ будет оптимальным. Однако строгого алгоритма для этого не существует, поэтому приходится действовать методом перебора: выбирают фиксированный шаг изменения λ (например, 0,01; 0,001; 0,0001) и для каждого λ

от 0 до 1 рассчитывают значения z_t. Для полученной модели вида (11.9) можно провести оценку параметров ($\alpha_0, \beta_0, \alpha_0, \beta_0, \lambda$), а также рассчитать коэффициент детерминации. Из построенных моделей выбирается та, которая обеспечивает наибольшее значение R^2.

Следует отметить, что количество лагов p задается таким образом, чтобы при увеличении их на единицу изменение z_t было бы меньше заранее выбранного малого числа Δ (например, если x_t описывает месячную заработную плату (в рублях), т.е. диапазоны ее изменения – от 500 до 10000, то Δ может быть порядка 1 – 10).

II. Преобразование Койка. Этот метод достаточно широко распространен в эконометрическом анализе. Он реализуется следующим образом: из уравнения (11.6) вычитается такое же уравнение, только рассчитанное для периода (*t-1*) и умноженное на λ:

$$\lambda \cdot y_{t-1} = \lambda \cdot \alpha_0 + \beta_0 \lambda^2 \cdot x_{t-1} + \beta_0 \lambda^3 \cdot x_{t-2} \ldots + \beta_0 \lambda^{p+1} \cdot x_{t-p-1} + \lambda \cdot \varepsilon_{t-1} \quad (11.10)$$

Получаем уравнение:

$$y_t - \lambda \cdot y_{t-1} = (1-\lambda) \cdot \alpha_0 + \beta_0 x_t + (\varepsilon_t - \lambda \cdot \varepsilon_{t-1}) \quad (11.11)$$

после несложных преобразований:

$$y_t = (1-\lambda) \cdot \alpha_0 + \beta_0 x_t + \lambda \cdot y_{t-1} + \nu_t \quad (11.12)$$

где $\nu_t = (\varepsilon_t - \lambda \cdot \varepsilon_{t-1})$ – скользящая средняя случайных остатков временного ряда – также случайная величина.

Преобразование модели вида (11.8) в модель вида (11.12) получила название *преобразования Койка*. Как видно из формулы, модель (11.12) является линейной регрессией с двумя объясняющими переменными и тремя неизвестными параметрами, оценить которые можно при помощи МНК.

Однако в этом случае возможны следующие проблемы, приводящие к несостоятельности и смещенности оценок, получаемых по МНК:

1. Переменная y_{t-1} носит случайный характер, и, кроме того, она, скорее всего, будет коррелировать с ν_t, что приводит к нарушению предпосылок МНК.
2. Для случайных отклонений ν_t вполне возможно наличие *автокорреляции* (см. ниже).

Для преодоления проблем, связанных с применением МНК, используют *обобщенный метод наименьших квадратов (ОМНК)*. Основное отличие ОМНК в том, что для его применения необязательно выполнение предпосылок 2 и 3 МНК, касающихся взаимонезависимости случайных остатков и постоянства их дисперсии. Применение ОМНК требует знания дополнительной информации об остатках, получить которую чрезвычайно сложно, поэтому на практике часто используют частные случаи ОМНК – *доступный метод наименьших квадратов* [Кремер, Путко, стр. 185] и *метод взвешенных наименьших квадратов*.

Модель (11.12) можно применять для *долгосрочного прогнозирования*. Если предположить, что x стремится к своему равновесному значению x^*, то:

$$y^* = (1-\lambda)\cdot\alpha_0 + \beta_0 x^* + \lambda\cdot y^* \Rightarrow y^* = \alpha_0 + \frac{\beta_0}{(1-\lambda)}x^* \qquad (11.13)$$

Следовательно, мы можем определить равновесное значение зависимой переменной.

Вообще для временных рядов и динамических регрессионных моделей достаточно распространено явление автокорреляции (взаимозависимости) случайных остатков или просто *автокорреляции*.

К *причинам возникновения автокорреляции* относят:

1. *Ошибки спецификации* – к примеру, если спецификация модели – линейная функция, а фактическая зависимость лучше описывается логарифмической или показательной функцией, возникает автокорреляция.

2. *Инерционность экономических законов*. Зачастую внешние по отношению к системе факторы воздействуют не моментно, а достаточно протяженно во времени. Экономическая система как бы адаптируется к внешним воздействиям, и после их исчезновения «по привычке» функционирует так, как будто это воздействие еще имеет место.

3. *Временные лаги в равновесных моделях.* Имеется в виду так называемый эффект паутины, когда равновесие устанавливается не сразу, а каждое следующее состояние изменяется как бы по спирали – последовательно отклоняясь от равновесного значения в разные стороны. В этом случае имеют дело с отрицательной автокорреляцией.

4. *Сглаживание данных.* Использование специальных статистических приемов для обработки исходных данных позволяет получить более удобный для анализа набор усредненных данных (к примеру, метод скользящего среднего). Однако в этом случае возникает взаимозависимость между соседними значениями, и как следствие, инерционность данных. В таком случае обнаруживаемая автокорреляция будет положительной.

Рассмотрим классификацию автокорреляции (рис. 11.3).

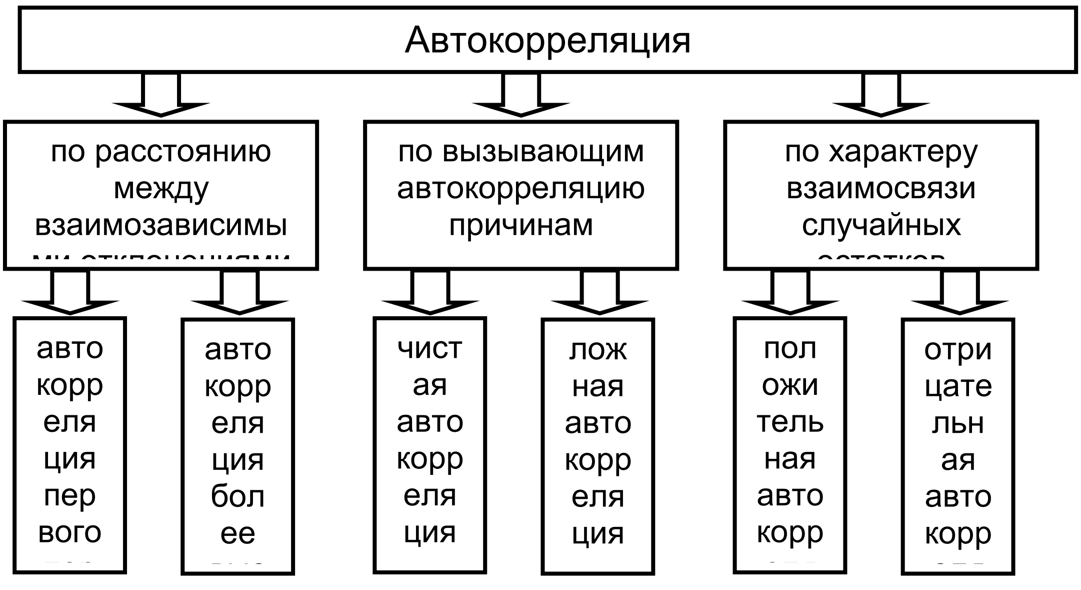


Рис. 11.3 Классификация автокорреляции

В зависимости от причин, вызывающих автокорреляцию, различают *чистую и ложную автокорреляцию*: чистая автокорреляция появляется как следствие зависимости случайного члена от прошлых значений (причины 2-4), а ложная – неверной спецификацией модели (причина 1).

В зависимости от того, между какими случайными отклонениями наблюдается взаимосвязь, различают порядки автокорреляции. Наиболее распространена *автокорреляция первого порядка* – автокорреляция между соседними случайными отклонениями. Вообще под *автокорреляцией i-го порядка* понимают зависимость между e_t и e_{t-i}. В дальнейшем мы будем говорить об автокорреляции первого порядка. Однако возможна и автокорреляция более высоких порядков. В частности, когда рассматриваются поквартальные данные с ярко выраженной сезонной составляющей, то, скорее всего, будет иметь место положительная автокорреляция четвертого порядка.

По характеру взаимозависимости случайных остатков различают *два вида автокорреляции* – положительную и отрицательную. При *положительной автокорреляции* для каждого наблюдения высока вероятность того, что $e_t e_t$ будет того же знака, что и $e_{t-1} e_{t-1}$. При *отрицательной автокорреляции* для большинства соседних отклонений знаки $e_t e_t$ и $e_{t-1} e_{t-1}$ не совпадают. Графически положительная и отрицательная автокорреляция выглядят следующим образом (рис. 11.4):

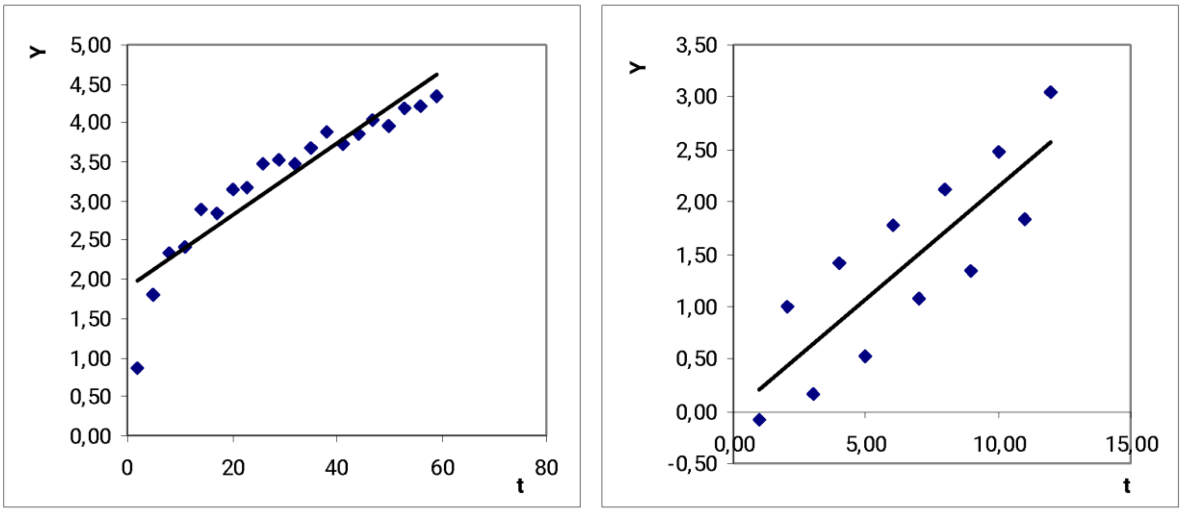


Рис. 11.4 Положительная (слева) и отрицательная (справа) автокорреляция

Рассмотрим причины возникновения, последствия, а также методы обнаружения и устранения автокорреляции (рис. 11.5).

Последствия автокорреляции:

1. Оценки неизвестных параметров неэффективны (однако остаются несмещенными).

2. Расчетные значения дисперсий оценок неизвестных параметров оказываются заниженными, что может привести к признанию статистической значимости незначимых в действительности параметров.

3. Ухудшаются прогнозные качества модели.

Рис. 11.5 Причины возникновения, последствия, методы обнаружения и устранения автокорреляции

Задания

1. На основе экономических знаний постройте динамические модели адаптивных и рациональных ожиданий. К каким типам относятся эти модели?

2. Рассматривается модель равновесия на рынке капусты. Уравнение кривой предложения имеет вид:

$$Q^s_t = -30 + 0,5P_{t-1} + 0,25P_{t-2} + \varepsilon_t$$

К какому типу относится эта модель?

3. Предположим, что в результате преобразования Койка и расчета параметров модель $DL(p)$ приобрела вид:

$$y_t = 0,12 - 0,4x_t + 0,75y_{t-1} + \varepsilon_t$$

где y_t – доля предприятий на рынке в период t, в долях единицы

x_t – темп прироста емкости рынка в период t, в долях единицы

Определите значения коэффициентов α_0, β_0, λ исходной модели. К какому равновесному значению стремиться доля фирмы на рынке, если прогнозируется среднегодовой темп прироста емкости рынка 7%?

Литература: [1, 2, 4].

ЛАБОРАТОРНАЯ РАБОТА 12

Специфика построения динамических регрессионных моделей

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: культуру мышления и - способы обобщения информации.

Уметь: анализировать социально значимые проблемы и процессы, происходящие в обществе, и прогнозировать возможное их развитие в будущем и уметь строить стандартные теоретические и эконометрические модели.

Владеть: культурой мышления и современными методами сбора, обработки и анализа информации.

Актуальность темы: определяется тем, что большинство эконометрических моделей разработаны для анализа стационарного ряда, поскольку методы его анализа значительно проще, а нестационарный ряд достаточно легко приводится к стационарному путем исключения основной временной тенденции (тренда). Для нас особый интерес представляют факторные динамические модели, в которых в качестве объясняющих переменных используются переменные, принадлежащие различным периодам времени.

Теоретическая часть

Методы обнаружения автокорреляции:

I. *Графический метод.* Достаточно просто автокорреляцию можно обнаружить, анализируя графики остатков. Выделяют два вида таких графиков: первый показывает зависимость от фактора времени произведения соседних значений случайных отклонений (рис. 12.6), второй – взаимосвязь между соседними отклонениями (рис. 12.7).

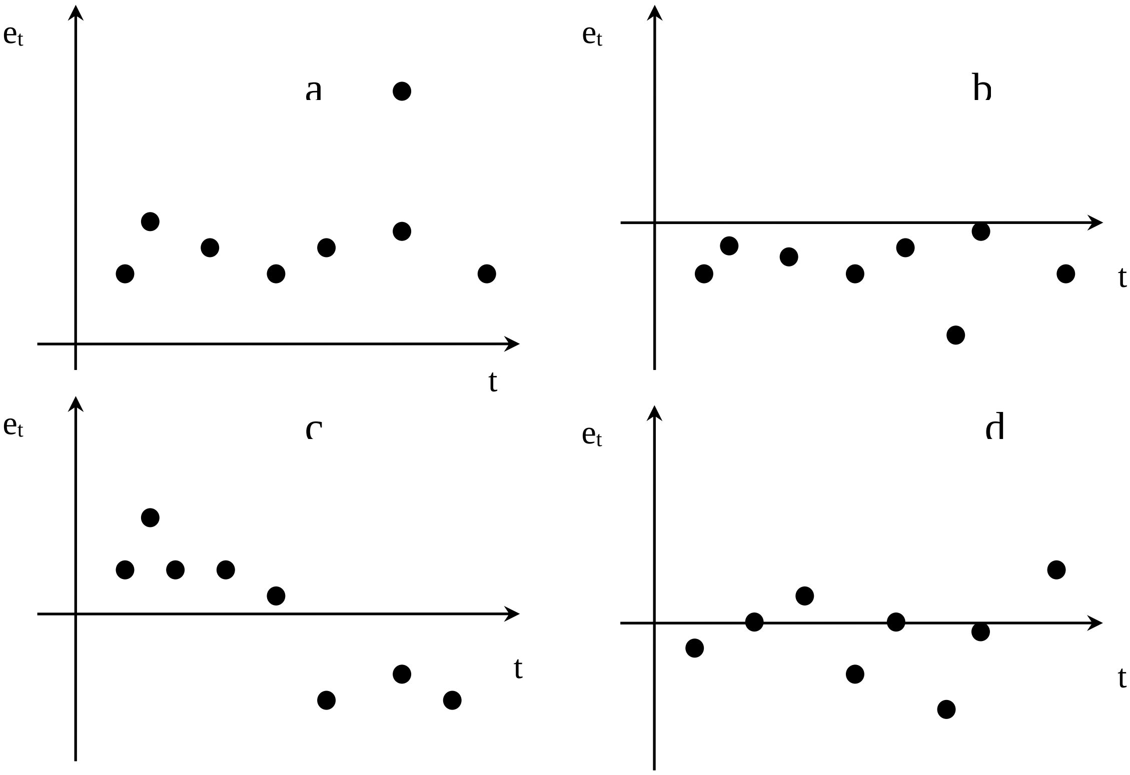


Рис. 12.6 Обнаружение автокорреляции графическим методом
(анализ динамики произведения остатков):
a) положительная автокорреляция; b) отрицательная автокорреляция; c) смешанная автокорреляция; d) автокорреляция отсутствует

Рис. 12.7 Обнаружение автокорреляции графическим методом
(анализ корреляционного поля соседних остатков)
a) положительная автокорреляция; b) отрицательная автокорреляция
c) отсутствие автокорреляции

Графические методы очень удобны, просты в применении, однако не позволяют точно установить, есть ли автокорреляция, в тех случаях, когда это не очевидно (рис. 12.8). Для них не существует однозначно определенного формального критерия, что сужает область их использования. Рекомендуется использовать графические методы для предварительной оценки наличия автокорреляции.

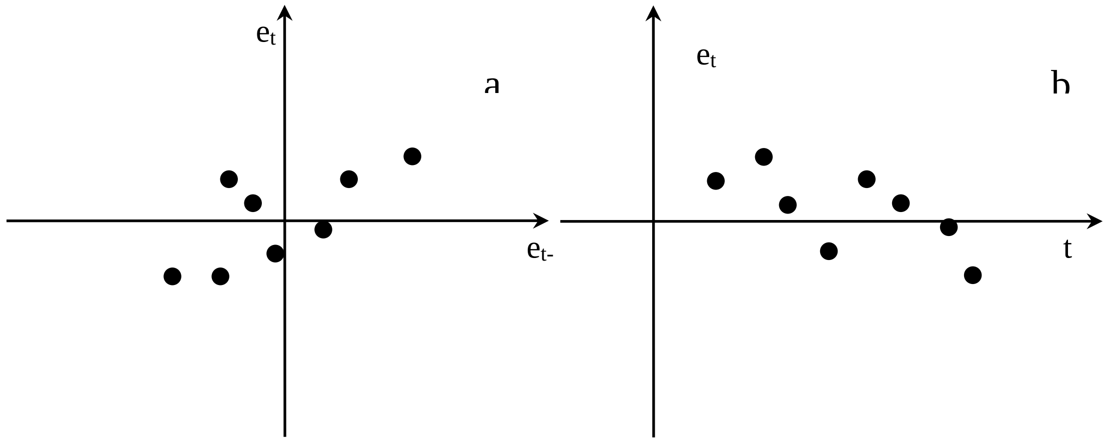


Рис. 12.8 Случаи, когда невозможно определить автокорреляцию при помощи графического метода

II. Метод рядов. Этот метод позволяет точно определить, существует ли автокорреляция в исследуемом наборе данных. Алгоритм применения метода состоит из следующих этапов:

1. Определяются знаки случайных отклонений, определяется количество рядов *k* – последовательности случайных отклонений с одним знаком. Для этого все случайные отклонения представляют в виде:

(– – – –)(+ + + + +)(– – – – – – – – – –)(+ +)(– – –)

2. Рассчитывают критические значения для *k* (k_{low} и k_{high}), исходя из предположения, что если автокорреляции нет, то *k* имеет нормальное распределение с математическим ожиданием *M(k)* и дисперсией *D(k)*:

$$M(k) = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2n_1n_2}{n} + 1 \tag{12.14}$$

$$D(k) = \frac{2n_1n_2 - n_1 - n_2}{(\mathrm{n}_1+\mathrm{n}_2)^2(\mathrm{n}_1+\mathrm{n}_2-2)} = \frac{2n_1n_2 - n}{n^2(\mathrm{n}-2)} \tag{12.15}$$

где: *n* – объем выборки

n_1 – общее количество единиц с положительным знаком случайного отклонения (количество знаков «+» в ряду)

n_2 – общее количество единиц с отрицательным знаком случайного отклонения (количество знаков «–» в ряду)

Для заданного уровня значимости α критические значения будут равны:

$$k_{low} = M(\mathrm{k}) - u_{\alpha/2}D(\mathrm{k}) \tag{12.16}$$

$$k_{high} = M(\mathrm{k}) - u_{\alpha/2}D(\mathrm{k}) \tag{12.17}$$

Сравнивается фактическое значение количества рядов k с критическими значениями k_{low} и k_{high}. Если k ≤ k_{low}, то делается вывод о наличии положительной автокорреляции остатков. Если $k \geq k_{high}$, то считается, что присутствует отрицательная автокорреляция. Если же $k_{low} < k < k_{high}$, то утверждается отсутствие автокорреляции.

Для небольшого числа наблюдений существуют специальные статистические таблицы k_{low} и k_{high}, разработанные Сведом и Эйзенхартом (см. приложение 5), при помощи которых упрощается процедура анализа автокорреляции.

III.Критерий Дарбина-Уотсона (DW)

Суть метода в анализе статистики Дарбина-Уотсона, рассчитываемой по формуле:

$$\mathrm{DW} = \frac{\sum_{t=2}^{T}(\mathrm{e}_t - \mathrm{e}_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \qquad (12.18)$$

Доказано, что статистика Дарбина-Уотсона связана с коэффициентом корреляции между соседними отклонениями по формуле:

$$DW \approx 2(1 - \mathrm{r}_{e_t e_{t-1}}) \qquad (12.19)$$

Следовательно, статистика DW принимает значения в интервале от 0 до 4. при этом можно сделать следующие выводы относительно значений DW (табл. 12.1)

Таблица 12.1 – Значения, принимаемые критерием Дарбина-Уотсона и их интерпретация

№п/п	Значения DW	Значения $\mathrm{r}_{e_t e_{t-1}}$	Вывод о наличии автокорреляции
1	близко к 0	близко к 1	сильная положительная автокорреляция
2	близко к 2	близко к 0	автокорреляция отсутствует
3	близко к 4	близко к –1	сильная отрицательная автокорреляция

Однако приведенной выше интерпретации статистики DW, как правило, недостаточно, чтобы сделать вывод о наличии либо отсутствии автокорреляции. Необходимо указать конкретные диапазоны DW, для которых характерно наличие или отсутствие автокорреляции. При грубой оценке предполагают, что автокорреляция не имеет место, когда 1,5 < DW < 2,5. Для точного определения наличия автокорреляции используются критические значения d_{low} и d_{high} из таблицы (приложение 4). Тогда интерпретация статистики DW примет вид:

Таблица 6.2 – Интерпретация значений критерия Дарбина-Уотсона с использованием критических значений

№п/п	Значения DW	Вывод о наличии автокорреляции
1	$[0;d_{low})$	существует положительная автокорреляция
2	$[d_{low};d_{high})$	нельзя сделать вывод о наличии автокорреляции
3	$[d_{high};4-d_{high})$	автокорреляция отсутствует
4	$[4-d_{high};4-d_{low})$	нельзя сделать вывод о наличии автокорреляции
5	$[4-d_{low};4]$	существует отрицательная автокорреляция

Однако наряду с очевидными достоинствами критерия Дарбина-Уотсона существуют и *ограничения* его применения на практике:

1. Критерий может применяться только для моделей, содержащих свободный член
2. Временной ряд, по которому построена модель, должен быть полным (то есть внутри ряда должны быть все данные).
3. Критерий Дарбина-Уотсона нельзя применять для авторегрессионных моделей $AR(p)$.
4. Критерий позволяет выявить автокорреляцию только первого порядка

Рассмотрим методы устранения автокорреляции. Выделяют два основных метода: при помощи первого устраняется сама автокорреляция, а при помощи второго возможно получение эффективных оценок коэффициентов и более точных значений их дисперсий. Рассмотрим каждый из методов:

1) *Изменение спецификации модели.* Если автокорреляция является следствием неверной спецификации модели, то ее можно устранить путем изменения спецификации. Обычно в модель может быть добавлена периодическая составляющая (в виде тригонометрической функции – рис. 12.9) или учтен нелинейный характер изменений путем замены функции на логарифмическую или показательную в зависимости от характера отклонений (рис. 12.10).

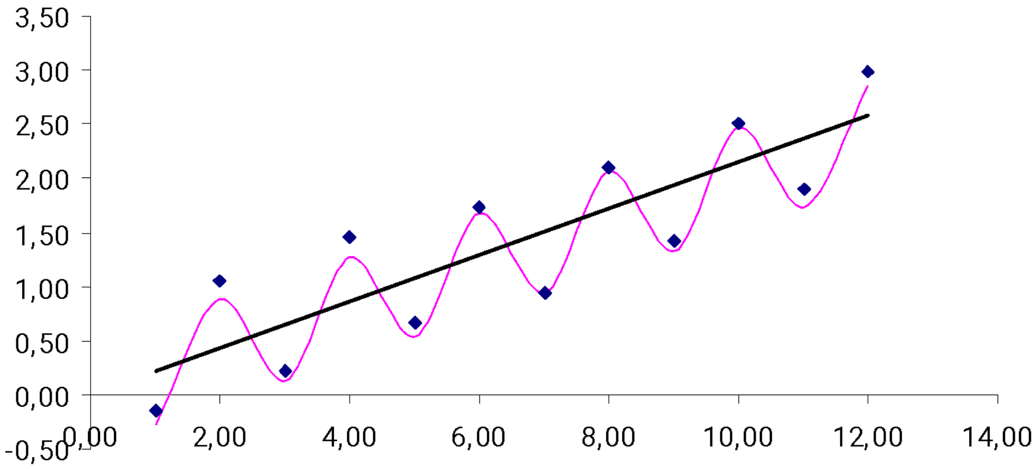


Рис. 12.9 Устранение отрицательной автокорреляции путем изменения спецификации модели с линейной функции на комбинированную (тригонометрическую-линейную)

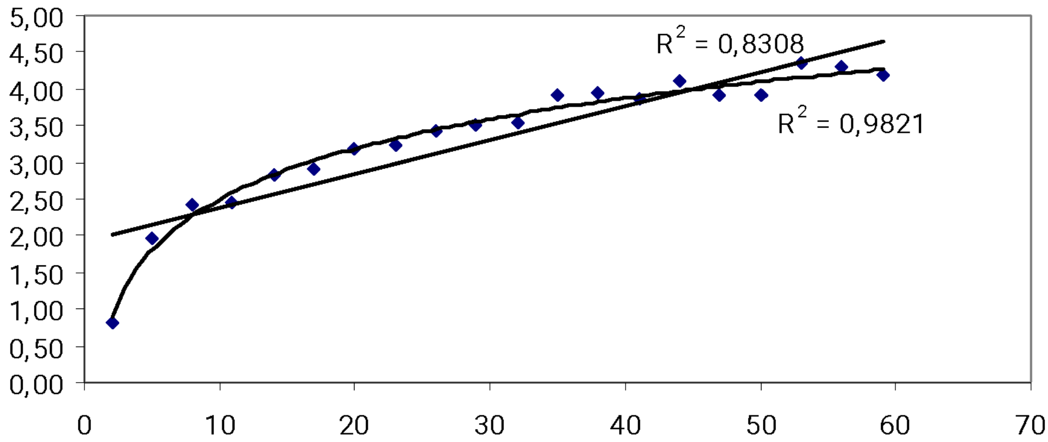


Рис. 12.10 Устранение автокорреляции изменением спецификации модели с линейной на логарифмическую

2) *Авторегрессионное преобразование.* Для выполнения преобразования воспользуемся алгоритмом:

1. Из исходной модели $y = \alpha_0 + \alpha_1 x + \varepsilon$, в которой наблюдается автокорреляция, строим модель регрессионной зависимости между случайными остатками $e_t = \rho e_{t-1} + u_t$, которая будет удовлетворять предпосылкам МНК.

2. Строим два уравнения для периодов *t* и *t-1*:

$$y_t = \alpha_0 + \alpha_1 x_t + \varepsilon_t \tag{12.20}$$

$$y_{t-1} = \alpha_0 + \alpha_1 x_{t-1} + \varepsilon_{t-1} \tag{12.21}$$

3. Вычитаем из первого уравнения второе, умноженное на ρ:

$$y_t - \rho \cdot y_{t-1} = (\alpha_0 - \rho \cdot \alpha_0) + (\alpha_1 \mathrm{x}_t - \rho \cdot \alpha_1 \mathrm{x}_{t-1}) + (\varepsilon_t - \rho \cdot \varepsilon_{t-1})$$

$$y_t - \rho \cdot y_{t-1} = \alpha_0 (1 - \rho) + \alpha_1 (\mathrm{x}_t - \rho \cdot \mathrm{x}_{t-1}) + (\varepsilon_t - \rho \cdot \varepsilon_{t-1})$$

4. Заменим: $\mathrm{y}_t^* = y_t - \rho \cdot y_{t-1}; x_t^* = x_t - \rho \cdot x_{t-1}; \mathrm{u}_t = \varepsilon_t - \rho \cdot \varepsilon_{t-1}; \alpha_0^* = \alpha_0 (1 - \rho)$, тогда уравнение примет вид:

$$\mathrm{y}_t^* = \alpha_0^* - \alpha_1 x_t^* + u_t \tag{12.22}$$

К уравнению модели (12.22) можно применить МНК, так как для него случайные остатки удовлетворяют условиям Гаусса-Маркова. В результате применения МНК получаем оценки параметров a_0* и a_1.

5. Зная параметр модели (6.17) a_0*, определяем параметр исходной модели (12.20) $a_0 = \dfrac{a_0^*}{1 - \rho}$

Авторегрессионное преобразование уменьшает число степеней свободы на 1, что при небольших объемах выборок приводит к снижению эффективности оценок. Это проблема решается при помощи *поправки Прайса-Винстена*:

$$x_1^* = \sqrt{1-\rho^2 x_1} \qquad (12.23)$$

$$y_1^* = \sqrt{1-\rho^2 y_1} \qquad (12.24)$$

Авторегрессионное преобразование применимо и в случае множественной регрессии.

Основная проблема, связанная с использованием авторегрессионного преобразования – оценка параметра ρ. Для этого используют следующие *методы*:

1. *Статистика Дарбина-Уотсона*. В этом случае предполагают, что при больших выборках $r_{e_t e_{t-1}} \approx \rho$. Тогда, используя соотношение $DW \approx 2(1-\mathrm{r}_{e_t e_{t-1}})$, получим выражение для оценки ρ:

$$\rho \approx 1-\frac{DW}{2} \qquad (12.25)$$

2. *Метод Кохрана-Оркатта*. Этот метод относится к числу итеративных и основан на применении следующего алгоритма:

1) По МНК определяются параметры в регрессии:

$$y = \alpha_0 + \alpha_1 x + \varepsilon \qquad (12.26)$$

и рассчитываются случайные ошибки $e_t = y_t - \hat{y}_t$

2) При помощи МНК оценивается параметр $\hat{\rho}$ в уравнении регрессии:

$$e_t = \hat{\rho} e_{t-1} + u_t \qquad (6.27)$$

3) Полученное значение $\hat{\rho}$ подставляется в уравнение, полученное при помощи авторегрессионного преобразования:

$$\mathrm{y}_t^* = \alpha_0(1-\hat{\rho}) + \alpha_1 \cdot x_t^* + u_t \qquad (12.28)$$

для которого по МНК оцениваются параметры α_0 и α_1.

4) Рассчитанные на предыдущем этапе оценки параметров a_0 и a_1 подставляются в уравнение (12.19), затем этапы 2 и 3 алгоритма повторяются до тех пор, пока разница между параметрами $\hat{\rho}$, полученными на последовательных этапах, не будет меньше заранее заданного числа (0,01; 0,001 и т.п.).

3. *Метод Хилдрета-Лу*.

С заданным шагом (0,01; 0,001 и т.п.) осуществляется перебор значений ρ из диапазона [–1;1]. Для каждой модели оценивается качество модели (12.17). Выбирается модель, обеспечивающая наилучшее качество.

Тесты

1. В какой модели учитываются лаги и в объясняющих, и в зависимой переменной?

а) AR(p)

b) ARMA (p,q)
c) ARIMA (p,q,n)
d) ADL (p,q)

2. Какой метод построения авторегрессионных моделей приводит к значительной потере числа степеней свободы?

a) метод геометрической прогрессии
b) метод последовательного увеличения количества лагов
c) метод моментов
d) метод наименьших квадратов

3. Преобразование Койка превращает:

a) модель AR(n) в модель ARMA(n,2)
b) модель MA(n) в модель ARMA(1,1)
c) модель AR(n) в модель ADL(1,1)
d) модель ADL(n,m) в модель ARMA(n,m)

4. Какие из условий Гаусса – Маркова снимаются при применении обобщенного метода наименьших квадратов (см. тему 3)?

a) 1 и 2
b) 1, 2, 3
c) 2 и 3
d) все

5. Под автокорреляцией понимают:

a) существование авторегрессионных моделей с числом лагов не менее 3
b) взаимозависимость случайных остатков
c) взаимозависимость независимых переменных
d) чередование знаков произведений случайных отклонений

Литература: [1, 2, 3].

ЛАБОРАТОРНАЯ РАБОТА 13

Гетероскедастичность в регрессионных моделях

Цель: формирование у студентов профессиональной компетенции

Врезультате освоения темы студент должен:

Знать: культуру мышления и - способы обобщения информации.

Уметь: анализировать социально значимые проблемы и процессы, происходящие в обществе, и прогнозировать возможное их развитие в будущем и уметь строить стандартные теоретические и эконометрические модели.

Владеть: культурой мышления и современными методами сбора, обработки и анализа информации.

Актуальность темы: определяется тем, что истинная гетеро- скедастичность возникает в пространственных выборках как следствие влияния на вариацию зависимой переменной фактора пропорциональности (одной из независимых переменных), во временных рядах, когда значения зависимой переменной качественно неоднородны (к примеру, действуют различные институциональные факторы) или для нее характерен высокий темп изменения, в том случае, если качество данных внутри выборки неоднородно.

Теоретическая часть

Гетероскедастичность – это различие в дисперсиях случайных отклонений при различных значениях зависимой переменной. Наличие гетероскедастичности фактически означает невыполнение одной из предпосылок применения МНК (условие постоянства дисперсий). Гетероскедастичность характерна в первую очередь для перекрестных данных (относящихся к одному моменту времени, но к разным единицам наблюдения).

По причине возникновения различают истинную и ложную гетероскедастичность. *Истинная гетероскедастичность* вызывается непостоянством дисперсии случайного члена и ее зависимостью от различных факторов.

Истинная гетероскедастичность возникает:

- в пространственных выборках как следствие влияния на вариацию зависимой переменной *фактора пропорциональности* (одной из независимых переменных).
- во временных рядах, когда значения зависимой переменной качественно неоднородны (к примеру, действуют различные институциональные факторы) или для нее характерен высокий темп изменения
- в том случае, если качество данных внутри выборки неоднородно.

Ложная гетероскедастичность возникает как следствие неправильной спецификации модели регрессии.

Наличие гетероскедастичности влечет за собой следующие *последствия:*

- оценки коэффициентов, полученные по МНК, становятся неэффективными;
- дисперсии коэффициентов будут заниженными относительно действительных значений, что может привести к признанию статистической значимости на самом деле незначимых коэффициентов регрессии. Кроме этого, построенные доверительные интервалы для зависимой переменной будут *у*же.

Рассмотрим *методы обнаружения* гетероскедастичности:

Графический анализ остатков. По результатам построения модели и расчета случайных остатков строят график на осях $(x_i;e_i^2)$ (рис. 13.2).

Причины гетероскедастичности:

влияние на вариацию зависимой переменной фактора пропорциональности

во временных рядах: качественная неоднородность значений зависимой переменной или высокий темп ее изменения

неоднородность качества данных внутри выборки

неправильная спецификация модели

Последствия гетероскедастичности:

неэффективность оценок

признание статистической значимости незначимых переменных

сужение доверительных интервалов относительно их действительных значений

Методы обнаружения гетероскедастичности:

графический анализ остатков

тест ранговой корреляции Спирмена

тест Парка

тест Глейзера

тест Голдфельда – Квандта

тест Уайта

Методы устранения гетероскедастичности:

метод взвешенных наименьших квадратов

дисперсия случайных отклонений прямопропорциональна значениям независимой переменной

дисперсия случайных отклонений прямопропорциональна квадратам значений независимой переменной

Рис. 13.1 Причины, последствия, методы обнаружения и устранения последствий гетероскедастичности

К достоинствам метода следует отнести простоту реализации и возможность выявления гетероскедастичности любой формы. В то же время при помощи этого метода нельзя количественно оценить наличие гетероскедастичности, вследствие чего возникают проблемы при практической реализации метода, к примеру, для наборов данных, представленных на рис. 13.3

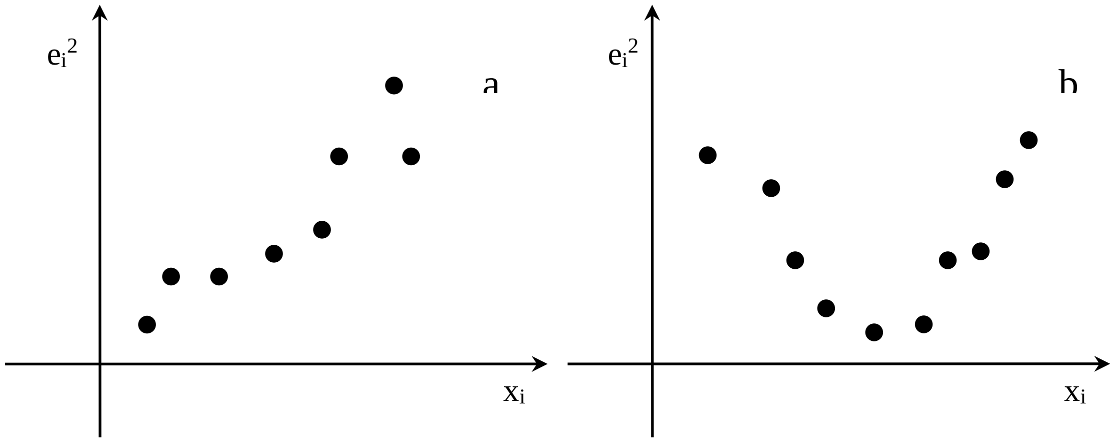


Рис. 13.2 Графический анализ остатков. Определение наличия гетероскедастичности: a – дисперсия остатков возрастает; b – дисперсия остатков сначала убывает, потом возрастает

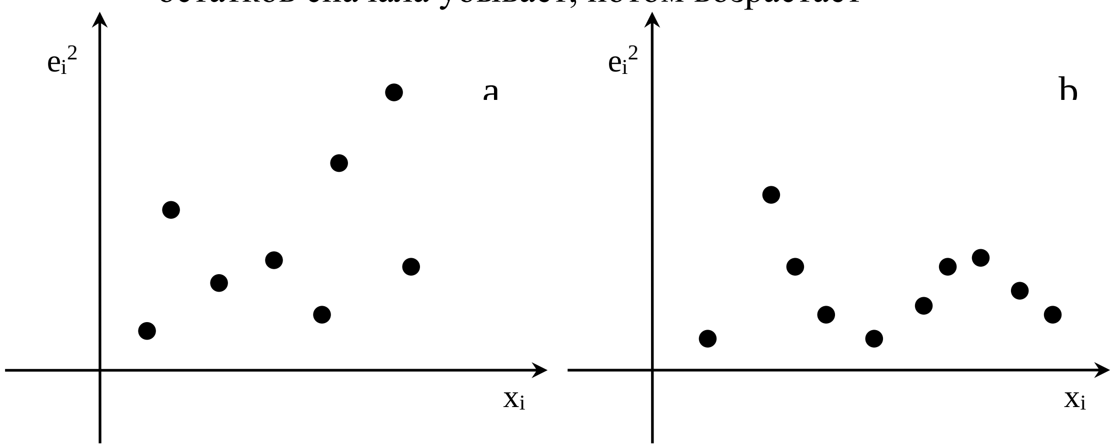


Рис. 13.3 Невозможность установления гетероскедастичности при помощи графического анализа остатков

1. *Тест ранговой корреляции Спирмена.* В этом случае рассчитывают коэффициент ранговой корреляции Спирмена между независимой переменной и случайными остатками:

$$\rho_{\text{x,e}} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{13.1}$$

где d_i – ранговая разность между рангами значений x_i и e^2_i.

Если связь оказывается статистически значимой, то делается вывод о наличии гетероскедастичности. Проверка статистической значимости связи осуществляется на основе *t-статистики*:

$$t = \frac{\rho_{x,e}\sqrt{n-2}}{\sqrt{1-\rho_{x,e}^2}} \tag{13.2}$$

Полученное значение сравнивают с табличным значением $t_{\alpha/2,n-2}$. Если расчетное значение оказывается больше, связь между независимой переменной и случайными остатками признается статистически значимой, и делается вывод о наличии гетероскедастичности.

В случае множественной регрессии тест ранговой корреляции проводится отдельно для каждой переменной.

К достоинствам метода следует отнести простоту реализации и возможность сделать точный вывод о наличии (или отсутствии) гетероскедастичности. Основной недостаток метода – его неприменимость в случаях, когда дисперсии остатков изменяются не монотонно (как на рис. 13.2 (б)).

2. *Тест Парка.* Этот тест дополняет метод графического анализа остатков и основан на построении регрессионной зависимости дисперсии остатков от значения зависимой переменной:

$$\sigma^2(\varepsilon_i) = \sigma^2 x_i^{\beta_i} e^{v_i} \qquad (13.3)$$

Модель (7.3) легко линеаризуется, и при помощи МНК можно получить оценку коэффициента β_1. Затем оценивается статистическая значимость этого коэффициента. Если коэффициент статистически значим, то наличие гетероскедастичности считается доказанным. В случае анализа множественной регрессии тест проводится отдельно для каждой переменной.

Формально тест Парка можно представить в виде следующего алгоритма (рис.13.4).

Достоинства теста Парка – существование жесткого критерия, позволяющего судить о наличии гетрероскедастичности. Недостаток – возможность необоснованных выводов, поскольку в тесте используется конкретная монотонно возрастающая (убывающая) функция. К примеру, применение теста к ситуации, изображенной на рисунке 13.1 (б), покажет, что гетероскедастичности нет.

3. *Тест Глейзера* по алгоритму применения аналогичен тесту Парка, однако за счет использования более универсальной функциональной зависимости он позволяет преодолеть недостаток теста Парка. В этом тесте используется функция:

$$\left|\varepsilon_{\mathrm{i}} = \beta_0 + \beta_1 x_i^k + u_i\right| \qquad (13.4)$$

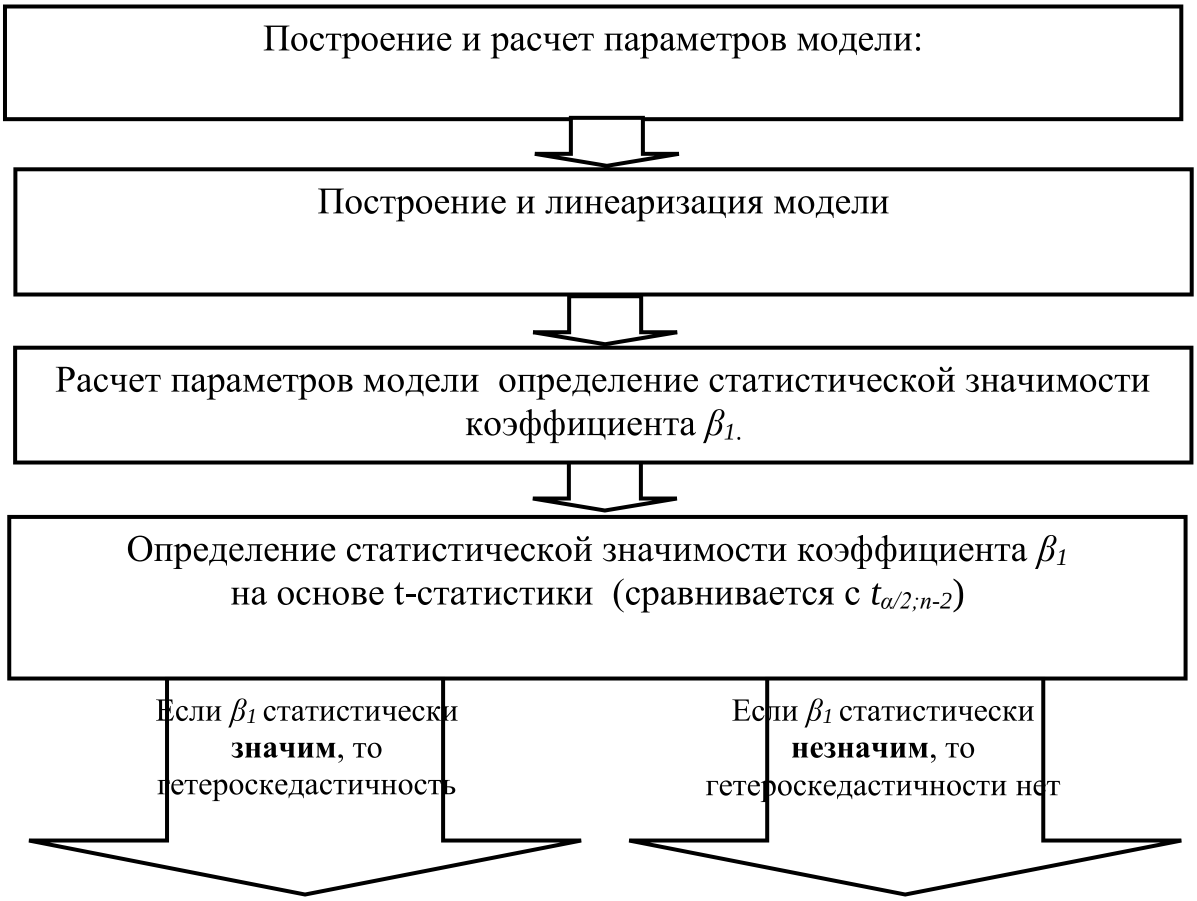


Рис. 13.4 Алгоритм применения теста Парка

где k может принимать любые значения.

Обычно используются следующие значения k: –1; –0,5; 0,5; 1; 2. Модель (13.4) строится для каждого значения k, оцениваются ее параметры и проверяется статистическая значимость β_1. Если хотя бы в одном случае этот коэффициент оказывается статистически значимым, то гетероскедастичность имеет место.

Тест Глейзера позволяет установить наличие гетероскедастичности даже в случаях немонотонного изменения функции $\varepsilon_i = f(x_i)$.

Однако и в тесте Парка, и в тесте Глейзера не учитывается возможность появления гетероскедастичности в регрессиях (13.3) и (13.4). Если она будет иметь место, то вероятность обнаружения гетероскедастичности там, где ее в действительности нет, возрастет.

Задания

1. Заполните таблицу:

№ п/п	Метод обнаружения гетероскедастичности	Достоинства	Недостатки
1.	…		
2.	…		
3.	…		

4.	…		
5.	…		

2. Определите, имеет ли место гетероскедастичность для следующих данных об объеме спроса на рынке бензина:

Цена (P), руб.	10	11,5	10,5	13	11	12	10,5	12	14	11,5	12,5	13	13	11,5	14
Объем спроса, (Q) тыс. л.	27	24	26	20	24	23	24	22	15	24	22	22	17	23	20

Для ответа используйте все известные Вам методы

3. В результате проведения маркетингового исследования ранка стирального порошка получены следующие результаты:

№ п/п	Среднемесячный объем потребления (Q), кг/мес.	Количество человек в семье *(N)*	Доход семьи *(I)*, тыс. руб.	Количество детей до 15 лет (N_d)
1.	3,4	1	2,7	0
2.	3,6	1	3,1	0
3.	5,0	2	3,8	0
4.	5,5	2	4,8	0
5.	6,7	2	5,2	1
6.	9,4	3	10,5	0
7.	10,7	3	11,1	1
8.	10,0	3	11,4	0
9.	11,2	3	12,3	1
10.	9,5	4	6,4	1
11.	12,6	4	10,8	2
12.	13,6	4	12,4	2
13.	13,1	4	14,0	1
14.	19,8	5	21,0	3
15.	19,9	7	16,8	3

Постройте регрессионную модель $Q = f(N,I,N_d)$

Определите наличие гетероскедастичности при помощи:

a) графического анализа остатков

b) теста Парка

c) теста Голдфельда-Квандта

Для каждой переменной оцените параметры модели при помощи обычного метода наименьших квадратов и метода взвешенных наименьших квадратов. Сравните полученные результаты.

Литература: [1, 2, 4].

ЛАБОРАТОРНАЯ РАБОТА 14

Гетероскедастичность в регрессионных моделях

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: культуру мышления и - способы обобщения информации.

Уметь: анализировать социально значимые проблемы и процессы, происходящие в обществе, и прогнозировать возможное их развитие в будущем и уметь строить стандартные теоретические и эконометрические модели.

Владеть: культурой мышления и современными методами сбора, обработки и анализа информации.

Актуальность темы: определяется тем, что истинная гетеро- скедастичность возникает в пространственных выборках как следствие влияния на вариацию зависимой переменной фактора пропорциональности (одной из независимых переменных), во временных рядах, когда значения зависимой переменной качественно неоднородны (к примеру, действуют различные институциональные факторы) или для нее характерен высокий темп изменения, в том случае, если качество данных внутри выборки неоднородно.

Теоретическая часть

Тест Голдфельда-Квандта. В этом тесте используется предположение, что дисперсия ошибок модели пропорциональна значениям независимой переменной. Исходная совокупность упорядочивается по мере возрастания значений независимой переменной, после чего исходная совокупность делится на 3 части размерами k, $n\text{-}2k$, k. Для первой и третьей подвыборок строят регрессионную модель, оценивают параметры и рассчитывают случайные отклонения. На основе этого определяется дисперсия случайных отклонений S^2_1 и S^2_3. Вывод о наличии гетероскедастичности делается в том случае, если различие между дисперсиями, определенное по критерию Фишера, оказывается статистически значимым. Формализовано тест Голдфельда-Квандта представим в виде следующего алгоритма (см. рис.14.1)

Рис.14.1 Алгоритм применения теста Голдфельда-Квандта

Рекомендуется выбирать значения k исходя из следующих соотношений:

Таблица 14.1 – Рекомендуемые значения k для выборок различных размеров

n	k
30	11
60	22

При применении этого теста для анализа гетероскедастичности в моделях множественной регрессии рекомендуется проводить его для той независимой переменной, значения которой в наибольшей степени связаны с дисперсией случайных остатков. Если такую переменную невозможно выделить, то тест проводят для всех переменных. Важно учитывать, чтобы значения k были больше чем общее значение переменных *(m+1)*.

Для смягчения проблемы гетероскедастичности используют *метод взвешенных наименьших квадратов (ВНК)*.

Основная идея метода состоит в том, что для смягчения проблемы гетероскедастичности следует «взвесить» наблюдаемые значения переменных с учетом соответствующей им дисперсии случайных отклонений (для этого используют среднеквадратическое отклонение ошибок: $\sigma_\varepsilon = \sqrt{\sigma_\varepsilon^2}$).

В этом случае исходная модель преобразуется:

$$\mathrm{y_i} = \alpha_0 + \alpha_1 x_i + \varepsilon_i \quad (14.1)$$

$$\frac{y_i}{\sigma_i} = \frac{\alpha_0}{\sigma_i} + \alpha_1 \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \quad (14.2)$$

введя замены: $y_i{*} = \frac{y_i}{\sigma_i}$; $z_i = \frac{1}{\sigma_i}$; $x_i{*} = \frac{x_i}{\sigma_i}$; $u_i = \frac{\varepsilon_i}{\sigma_i}$ получаем модель, для которой выполняются предпосылки МНК:

$$y_i{*} = \alpha_0 z_i + \alpha_1 x_i{*} + u_i \quad (14.3)$$

Следовательно, рассчитанные по МНК оценки будут обладать свойством эффективности. Единственной проблемой, препятствующей реализации ВНК, является то, что обычно дисперсия случайных отклонений является величиной неизвестной.

Для преодоления этой проблемы используют два способа оценивания дисперсий: в первом случае предполагают, что они прямопропорциональны значениям независимой переменной, во втором – прямопропорциональны значениям квадратов независимых переменных.

Если предполагается, что дисперсии случайных отклонений прямопропорциональны значениям объясняющей переменной, то при анализе парной регрессии преобразованное уравнение регрессии (14.2) примет вид:

$$\frac{\mathrm{y_i}}{\sqrt{\mathrm{x_i}}} = \frac{\alpha_0}{\sqrt{\mathrm{x_i}}} + \alpha_i \frac{\mathrm{x_i}}{\sqrt{\mathrm{x_i}}} + \frac{\varepsilon_\mathrm{i}}{\sqrt{\mathrm{x_i}}} \quad (14.4)$$

Когда устраняется гетероседастичность в моделях множественной регрессии, обычно в качестве весов используют полученные теоретические значения зависимой переменной

$$y{*} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m \quad (14.5)$$

и модель (7.6) имеет вид:

$$\frac{y_i}{\sqrt{y_i{*}}} = \frac{\alpha_0}{\sqrt{y_i{*}}} + \frac{x_i}{\sqrt{y_i{*}}} + \frac{\varepsilon_i}{\sqrt{y_i{*}}} \quad (14.6)$$

Если же используется предположение о пропорциональности дисперсий ошибок квадратам значений индивидуальных переменных, то модель (7.6) будет иметь вид:

$$\frac{\mathrm{y_i}}{\mathrm{x_i}} = \frac{\alpha_0}{\mathrm{x_i}} + \alpha_1 + \frac{\varepsilon_\mathrm{i}}{\mathrm{x_i}} \quad (14.7)$$

Следует отметить несколько общих положений, связанных с проблемой гетероскедастичности.

1. Существуют ситуации, когда гетероскедастичность обусловлена влиянием переменной, не включенной в модель. В этом случае следует эту переменную отыскать и включить в анализируемую регрессию.

На практике следует применять параллельно несколько методов обнаружения и преодоления гетероскедастичности.

Тесты

1. Антоним слова гетероскедастичность:

a) автокорреляция

b) гомоскедастичность

c) регрессия

d) различие в дисперсиях

2. При проведении теста Парка для установления наличия гетероскедастичности проверяется статистическая значимость:

a) коэффициента α_0 регрессии $y = \alpha_0 + \alpha_1 x + \varepsilon$

b) коэффициента α_1 регрессии $y = \alpha_0 + \alpha_1 x + \varepsilon$

c) коэффициента β_1 регрессии $\varepsilon_i^2 = \sigma^2 x^{\beta 1} e^{\nu 1}$

d) коэффициента σ регрессии $\varepsilon_i^2 = \sigma^2 x^{\beta 1} e^{\nu 1}$

3. Какой из тестов позволит, скорее всего, определить гетероскедастичность в следующем случае:

a) тест ранговой корреляции Спирмена

b) тест Парка

c) тест Глейзера

d) тест Голдфельда-Квандта

4. При проведении какого теста неважно, существует ли гетероскедастичность в регрессии $|\varepsilon_i| = f(x_i)$?

a) Тест Парка

b) Тест Глейзера при К = -1

c) Тест Глейзера при К = 1

d) Тест Голдфельда-Квандта

5. Если в результате «взвешивания» линейной модели получилось уравнение $\frac{y_i}{x_i} = \frac{a_0}{x_1} + a_1 + \frac{\varepsilon_1}{x_i}$, то использовалось предположение, что среднеквадратическое отклонение ошибок пропорционально:

a) x_i

b) $1/x_i$
c) $2x_i$
d) x_i^2

Литература: [1, 2, 3].

ЛАБОРАТОРНАЯ РАБОТА 15

Эконометрические модели в виде систем одновременных уравнений

Цель: формирование у студентов профессиональной компетенции ПК

В результате освоения темы студент должен:

Знать: пути анализа информации и возможности для постановки и решения целей.

Уметь: собирать необходимые данные, анализировать их и подготавливать информационные обзоры или аналитические отчеты и саморазвивать- ся повышая свою квалификацию и мастерство.

Владеть: навыками работы с компьютером как средством управления информацией.

Актуальность темы: определяется тем, что система линейных одновременных уравнений состоит из следующих компонентов системы: приведенная форма уравнений, косвенный метод наименьших квадратов, идентифицируемость системы уравнений, двухшаговый метод наименьших квадратов, трехшаговый метод наименьших квадратов.

Теоретическая часть

Проведение идентификации системы. Нахождение параметров уравнений системы.

Изучается модель вида

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1 \\ I_t = a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2 \\ r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3 \\ Y_t = C_t + I_t + G_t \end{cases}$$

где C_t – расходы на потребление в период t, Y_t – совокупный доход в период t, I_t – инвестиции в период t, r_t – процентная ставка в период t, M_t – денежная масса в период t, G_t – государственные расходы в период t, C_{t-1} – расходы на потребление в период $t-1$, I_{t-1} инвестиции в период $t-1$. Первое уравнение – функция потребления, второе уравнение – функция инвестиций, третье уравнение – функция денежного рынка, четвертое уравнение – тождество дохода.

Модель представляет собой систему одновременных уравнений. Проверим каждое ее уравнение на идентификацию.

Модель включает четыре эндогенные переменные (C_t,I_t,Y_t,r_t) и четыре предопределенные переменные (две экзогенные переменные – M_t и G_t и две лаговые переменные – C_{t-1} и I_{t-1}).

Проверим необходимое условие идентификации для каждого из уравнений модели.

Первое уравнение: $\mathrm{C_t} = a_1 + b_{11}Y_t + b_{12}C_{t-1} = \varepsilon_1$. Это уравнение содержит две эндогенные переменные C_t и Y_t и одну предопределенную переменную C_{t-1}. Таким образом, $H=2$, а $D=2+1=3$, т.е. выполняется условие $D+1>H$. Уравнение сверхидентифицируемо.

Второе уравнение: $I_t = a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2$. Оно включает две эндогенные переменные I_t и r_t и одну экзогенную переменную I_{t-1}. Выполняется условие $D+1=3+1>H=2$. Уравнение сверхидентифицируемо.

Третье уравнение: $r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon$. Оно включает две эндогенные переменные Y_t и r_t и одну экзогенную переменную M_t. Выполняется условие $D+1=3+1>H=2$. Уравнение сверхидентифицируемо.

Четвертое уравнение: $Y_t = C_t + I_t + G_t$. Оно представляет собой тождество, параметры которого известны. Необходимости в идентификации нет.

Задания

1. Постройте модели систем одновременных уравнений для различных экономических проблем.

№ п/п	Суть проблемы	Исходная система			Уравнения в приведенной форме	Характеристика идентифицируемости
		Система	Экзогенные переменные	Эндогенные переменные		

1.	Рыночное равновесие					

2. По приведенным в таблице данным постройте модель, охарактеризуйте ее, рассчитайте параметры. Какие цена и объем установятся на рынке, если издержки фирмы будут составлять 25 руб./ед., а средний доход потребителей будет равен 2520 руб. При оценке используется любой из подходящих методов.

Таблица 1 – Исходные данные

Объем (Q)	Цена (P)	Доход потребителей (I)	Средние общие издержки (ATC)
443	4,5	2 196	25,3
466	4,0	2 018	24,2
464	4,8	2 367	25,9
464	4,6	2 256	25,6
482	5,6	2 764	25,8
531	5,8	2 890	25,4
454	5,0	2 459	26,0
483	4,7	2 291	25,7
505	5,8	2 857	25,5
514	5,2	2 539	26,0
540	6,1	3 067	24,7
486	5,2	2 564	26,0
507	5,2	2 543	26,0
474	5,3	2 623	26,0
507	6,0	3 016	24,9

Тесты

1. Внешние по отношению к модели переменные называются:
a) эндогенными
b) экзогенными
c) предопределенными
d) косвенными

2. Какими должны быть уравнения модели, чтобы к ним можно было применить косвенный метод наименьших квадратов?

a) структурными
b) поведенческими
c) линейными
d) приведенными

Литература: [1, 2, 3].

ЛАБОРАТОРНАЯ РАБОТА 16

Эконометрические модели в виде систем одновременных уравнений

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: пути анализа информации и возможности для постановки и решения целей.

Уметь: собирать необходимые данные, анализировать их и подготавливать информационные обзоры или аналитические отчеты и саморазвивать- ся повышая свою квалификацию и мастерство.

Владеть: навыками работы с компьютером как средством управления информацией.

Актуальность темы: определяется тем, что система линейных одновременных уравнений состоит из следующих компонентов системы: приведенная форма уравнений, косвенный метод наименьших квадратов, идентифицируемость системы уравнений, двухшаговый метод наименьших квадратов, трехшаговый метод наименьших квадратов.

Теоретическая часть

Проверим для каждого уравнения достаточное условие идентификации. Для этого составим матрицу коэффициентов при переменных модели.

	C_t	I_t	r_t	Y_t	C_{t-1}	I_{t-1}	M_t	G_t
I уравнение	–1	0	0	b_{11}	b_{12}	0	0	0
II уравнение	0	–1	b_{21}	0	0	b_{22}	0	0
III уравнение	0	0	–1	b_{31}	0	0	b_{32}	0

Тождество	1	1	0	–1	0	0	0	1

В соответствии с достаточным условием идентификации ранг матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, должен быть равен числу эндогенных переменных модели без одного.

Первое уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	I_t	r_t	I_{t-1}	M_t	G_t
II уравнение	–1	b_{21}	b_{22}	0	0
III уравнение	0	–1	0	b_{32}	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{22} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{22} b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Второе уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	Y_t	C_{t-1}	M_t	G_t
I уравнение	–1	b_{11}	b_{12}	0	0
III уравнение	0	b_{31}	0	b_{32}	0
Тождество	1	–1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12} b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Третье уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	I_t	C_{t-1}	I_{t-1}	G_t
I уравнение	–1	0	b_{12}	0	0
II уравнение	0	–1	0	b_{22}	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12} b_{22} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицируемы. Приведенная форма модели в общем виде будет выглядеть следующим образом:

$$\begin{cases} \mathrm{C_t} = A_1 + \delta_{11} C_{t-1} + \delta_{12} I_{t-1} + \delta_{13} M_t + \delta_{14} G_t + u_1 \\ I_t = A_2 + \delta_{21} C_{t-1} + \delta_{22} I_{t-1} + \delta_{23} M_t + \delta_{24} G_t + u_2 \\ r_t = A_3 + \delta_{31} C_{t-1} + \delta_{32} I_{t-1} + \delta_{33} M_t + \delta_{34} G_t + u_3 \\ Y_t = A_4 + \delta_{41} C_{t-1} + \delta_{42} I_{t-1} + \delta_{43} M_t + \delta_{44} G_t + u_4 \end{cases}$$

Задания

1. Проведите сравнительный анализ методов оценки систем одновременных уравнений. Результат занесите в таблицу.

№ п/п	Метод	Предпосылки	Ограничения
1.	…		
2.	…		
3.	…		

2. Постройте кейнсианскую модель равновесия на товарном рынке, если известно:

a) в экономике существуют налоги на потребление

b) величина инвестиций зависит от ВВП прошлого года и ставки процента

c) величина чистого экспорта зависит от курса доллара и ВВП.

Какие методы будут экзогенными, а какие эндогенными? Запишите построенную модель в приведенной форме. Охарактеризуйте идентифицируемость модели.

Тесты

1. Какие из свойств оценок МНК не выполняются при оценке параметров систем одновременных уравнений:

1. эффективность
2. несостоятельность
3. состоятельность

a) 1 и 2

b) 1 и 3
c) 2 и 3
d) 2

2. Если по коэффициентам приведенных уравнений можно получить несколько значений коэффициента исходных уравнений, то такая система называется:

a) идентифицируемой
b) неидентифицируемой
c) сверхидентифицируемой
d) линейной

3. Пусть n_u – количество коэффициентов исходных уравнений, n_y – количество уравнений для их определения. Система называется сверхидентифицируемой, если:

a) $n_u > n_y$
b) $n_u < n_y$
c) $n_u = n_y$
d) $n_u \neq n_y$

Литература: [1, 4].

ЛАБОРАТОРНАЯ РАБОТА 17

Моделирование одномерных временных рядов

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: пути анализа информации и возможности для постановки и решения целей.

Уметь: собирать необходимые данные, анализировать их и подготавливать информационные обзоры или аналитические отчеты и саморазвивать- ся повышая свою квалификацию и мастерство.

Владеть: навыками работы с компьютером как средством управления информацией.

Актуальность темы: определяется тем, что необходима интерпретация результатов решения для формирования необходимых представлений о возможностях использования трендовых моделей в экономическом анализе.

Теоретическая часть

Пусть имеются условные данные о средних расходах на конечное потребление (y_t, денежных единиц) за 8 лет (таблица 17.1).

Таблица 17.1

t	y_t	y_{t-1}	$y_t - \overline{y_1}$	$y_{t-1} - \overline{y_2}$	$(y_t - \overline{y_1})(y_{t-1} - \overline{y_2})$	$(y_t - \overline{y_1})^2$	$(y_{t-1} - \overline{y_2})^2$
1	7	-	-	-	-	-	-
2	8	7	-3,39	-3	9,87	10,8241	9
3	8	8	-3,29	-2	6,58	10,8241	4
4	10	8	-1,29	-2	2,58	1,6641	4
5	11	10	-0,29	0	0,00	0,0841	0
6	12	11	0,71	1	0,71	0,5041	1
7	14	12	2,71	2	5,42	7,3441	4
8	16	14	4,71	4	18,84	22,1841	16
Σ	86	70	-0,03	0	44,0	53,4287	38

По формулам

$$\overline{y_1} = \frac{\sum_{t=2}^{n} y_t}{n-1}; \quad \overline{y_2} = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1}$$

вычисляем

$$\overline{y_1} = \frac{8+8+10+11+12+14+16}{7} = 11{,}29,$$

$$\overline{y_2} = \frac{7+8+8+10+11+12+14}{7} = 10{,}0.$$

Далее, заполняем таблицу и используя формулу для вычисления линейного коэффициента корреляции, получаем

$$\mathrm{r}_1 = \frac{\sum_{t=2}^{n}(y_t - \overline{y_1})(y_{t-1} - \overline{y_2})}{\sqrt{\sum_{t=2}^{n}(y_t - \overline{y_1})^2 (y_{t-1} - \overline{y_2})^2}} = \frac{44}{\sqrt{53{,}43 * 38}} = 0{,}976$$

Полученное значение свидетельствует об очень тесной зависимостью между расходами на конечное потребление текущего непосредственно предшествующего годов и, следовательно, о наличии во временном ряде расходов на конечное потребление сильной линейной тенденции.

Задания

Имеются следующие данные об уровне безработицы (%) y_t за 8 месяцев:

Месяц	1	2	3	4	5	6	7	8
y_t	8,8	8,6	8,4	8,1	7,9	7,6	7,4	7,0

Задание:

1. Определите коэффициенты автокорреляции уровней этого ряда первого и второго порядка.

Тесты

1. Внешние по отношению к модели переменные называются:

a) эндогенными
b) экзогенными
c) предопределенными
d) косвенными

2. Какими должны быть уравнения модели, чтобы к ним можно было применить косвенный метод наименьших квадратов?

a) структурными
b) поведенческими
c) линейными
d) приведенными

Литература: [1, 2, 5].

ЛАБОРАТОРНАЯ РАБОТА 18

Моделирование одномерных временных рядов

Цель: формирование у студентов профессиональной компетенции

В результате освоения темы студент должен:

Знать: пути анализа информации и возможности для постановки и решения целей.

Уметь: собирать необходимые данные, анализировать их и подготавливать информационные обзоры или аналитические отчеты и саморазвиваться повышая свою квалификацию и мастерство.

Владеть: навыками работы с компьютером как средством управления информацией.

Актуальность темы: определяется тем, что необходима интерпретация результатов решения для формирования необходимых представлений о возможностях использования трендовых моделей в экономическом анализе.

Теоретическая часть

Пусть имеются некоторые условные данные об общем количестве правонарушений на таможне одного из субъектов (таблица 1).

Таблица 1

Год	Квартал	t	Количество возбужденных дел, y_t
2015	I	1	375
	II	2	371
	III	3	869

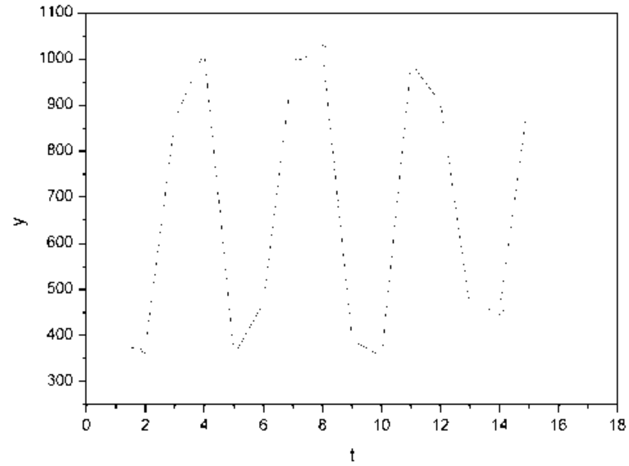


Рис. 18.1

	IV	4	1015
2016	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2017	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2018	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Построим поле корреляции:

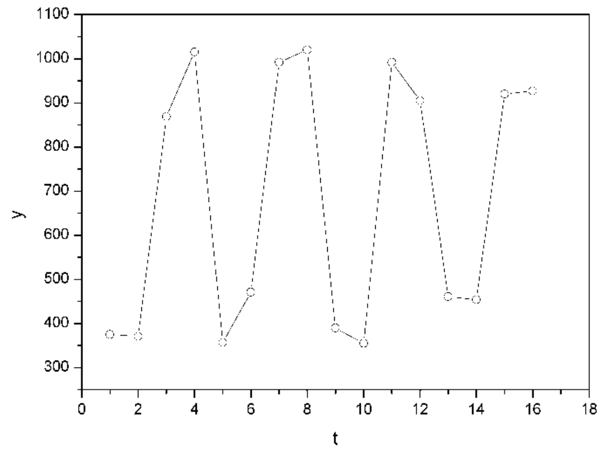


Рис. 18.1

Уже исходя из графика (рис. 18.1) видно, что значения y образуют пилообразную фигуру. Рассчитаем несколько последовательных коэффициентов автокорреляции. Для этого составляем первую вспомогательную таблицу (таблица 18.2).

Таблица 18.2

t	y_t	y_{t-1}	$y_t - \overline{y_1}$	$y_{t-1} - \overline{y_2}$	$(y_t - \overline{y_1})(y_{t-1} - \overline{y_2})$	$(y_t - \overline{y_1})^2$	$(y_{t-1} - \overline{y_2})^2$
1	2	3	4	5	6	7	8
1	375	–	–	–	–	–	–
2	371	375	-328,33	-288,13	94601,72	107800,59	83018,90
3	869	371	169,67	-292,13	-49565,70	28787,91	85339,94
4	1015	869	315,67	205,87	64986,98	99647,55	42382,46
5	357	1015	-342,33	351,87	-120455,66	117189,83	123812,50

6	471	357	-228,33	-306,13	69898,66	52134,59	93715,58
7	992	471	292,67	-192,13	-56230,69	85655,73	36913,94
8	1020	992	320,67	328,87	105458,74	102829,25	108155,48
9	390	1020	-309,33	356,87	-110390,60	95685,05	127356,20
10	355	390	-344,33	-273,13	94046,85	118563,15	74600,00
11	992	355	292,67	-308,13	-90180,41	85655,73	94944,10
12	905	992	205,67	328,87	67638,69	42300,15	108155,48
13	461	905	-238,33	241,87	-57644,88	56801,19	58501,10
14	454	461	-245,33	-202,13	49588,55	60186,81	40856,54
15	920	454	220,67	-209,13	-46148,72	48695,25	43735,36
16	927	920	227,67	256,87	58481,59	51833,63	65982,20
Сумма	10499	9947	9,05	0,05	74085,16	1153766,39	1187469,73
Среднее значение	699,33	663,13	–	–	–	–	–

Следует заметить, что среднее значение получается путем деления не на 16, а на 15, т.к. у нас теперь на одно наблюдение меньше.

Теперь вычисляем коэффициент автокорреляции первого порядка по формуле:

$$r_1 = \frac{74085{,}16}{\sqrt{115356{,}39 * 1187459{,}73}} = 0{,}063294$$

.

Составляем вспомогательную таблицу для расчета коэффициента автокорреляции второго порядка (таблица 18.3).

Таблица 18.3

t	y_t	y_{t-2}	$y_t - \overline{y_3}$	$y_{t-2} - \overline{y_4}$	$(y_t - \overline{y_3})(y_{t-2} - \overline{y_4})$	$(y_t - \overline{y_3})^2$	$(y_{t-2} - \overline{y_4})^2$
1	2	3	4	5	6	7	8
1	375	–	–	–	–	–	–
2	371	–	–	–	–	–	–
3	869	375	145,57	-269,79	-39273,33	21190,62	72786,64
4	1015	371	291,57	-273,79	-79828,95	85013,06	74960,96

5	357	869	-366,43	224,21	-82157,27	134270,94	50270,12
6	471	1015	-252,43	370,21	-93452,11	63720,90	137055,44
7	992	357	268,57	-287,79	-77291,76	72129,84	82823,08
8	1020	471	296,57	-173,79	-51540,90	87953,76	30202,96
9	390	992	-333,43	347,21	-115770,23	111175,56	120554,78
10	355	1020	-368,43	375,21	-138238,62	135740,66	140782,54
11	992	390	268,57	-254,79	-68428,95	72129,84	64917,94
12	905	355	181,57	-289,79	-52617,17	32967,66	83978,24
13	461	992	-262,43	347,21	-91118,32	68869,50	120554,78
14	454	905	-269,43	260,21	-70108,38	72592,52	67709,24
15	920	461	196,57	-183,79	-36127,60	38639,76	33778,76
16	927	454	203,57	-190,79	-38839,12	41440,74	36400,82
Сумма	10128	9027	-0,02	-0,06	-1034792,71	1037835,43	1116776,36
Среднее значение	723,43	644,79	–	–	–	–	–

Следовательно

$$r_2 = \frac{-1034792,71}{\sqrt{1037835,43 \cdot 1116776,36}} = -0,961183$$

.

$$r_1 = \frac{-1034792,71}{\sqrt{1037835,43 * 1116776,36}} = -0,961183 .$$

Аналогично находим коэффициенты автокорреляции более высоких порядков, а все полученные значения заносим в сводную таблицу 18.4.

Таблица 18.4

Лаг	Коэффициент автокорреляции уровней
1	0,063294
2	–0,961183
3	–0,036290
4	0,964735
5	0,050594

Анализ автокорреляционной функции и графика исходных уровней временного ряда позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала.

Задания

Имеются следующие данные об уровне безработицы (%) y_t за 8 месяцев:

Месяц	1	2	3	4	5	6	7	8
y_t	8,8	8,6	8,4	8,1	7,9	7,6	7,4	7,0

Задание:

2. Обоснуйте выбор уравнения тренда и определите его параметры.
3. Интерпретируйте полученные результаты.

Тесты

1. Какие из свойств оценок МНК не выполняются при оценке параметров систем одновременных уравнений:

1. эффективность
2. несостоятельность
3. состоятельность

a) 1 и 2
b) 1 и 3
c) 2 и 3
d) 2

2. Если по коэффициентам приведенных уравнений можно получить несколько значений коэффициента исходных уравнений, то такая система называется:

a) идентифицируемой
b) неидентифицируемой
c) сверхидентифицируемой
d) линейной
e) Пусть n_u – количество коэффициентов исходных уравнений, n_y – количество уравнений для их определения. Система называется сверхидентифицируемой, если:

a) $n_u > n_y$
b) $n_u < n_y$
c) $n_u = n_y$
d) $n_u \neq n_y$

Литература: [1, 2, 3].

Список рекомендованной литературы

Перечень основной литературы:

1. Афонин, В. В. Моделирование систем : учебное пособие / В. В. Афонин, С. А. Федосин. - 4-е изд. - Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2024. - 269 с. - ISBN 978-5-4497-2413-7. - Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. - URL: https://www.iprbookshop.ru/133951.html. Книга находится в премиум-версии ЭБС IPR BOOKS.

2. Яроцкая, Е. В. Экономико-математические методы и моделирование : учебное пособие / Е. В. Яроцкая. - 2-е изд. - Москва : Ай Пи Ар Медиа, 2026 . - 196 с. - ISBN 978-5-4497-3855-4. - Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. - URL: https://www.iprbookshop.ru/145188.html. Книга находится в премиум-версии ЭБС IPR BOOKS.

Перечень дополнительной литературы:

1. Наумов, И.В.; Эконометрика. Экономическое моделирование социально-экономических процессов в территориальных системах: учебное пособие / И. В. Наумов, Н. Л. Никулина. - Эконометрика. Экономическое моделирование социально-экономических процессов в территориальных системах, Весь срок охраны авторского права. - Электрон. дан. (1 файл). - Москва : Ай Пи Ар Медиа, 2021. - 127 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4497-1408-4,

2. Орлов, А. И. Эконометрика : учебное пособие / А. И. Орлов. - Эконометрика,2021-12-05. - Электрон. дан. (1 файл). - Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. - 676 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4497-0362-0,

3. Лихтенштейн, В.Е.; Математическое моделирование экономической справедливости Электронный ресурс : монография / Г.В. Росс / В.Е. Лихтенштейн. - Саратов : Ай Пи Эр Медиа, 2018. - 129 с. - Книга находится в базовой версии ЭБС IPRbooks. - ISBN 978-5-4486-0350-1

4. Петров, А. Е.; Математические модели принятия решений Электронный ресурс : Учебно-методическое пособие / А. Е. Петров. - модели принятия решений,2019-09-01. - Москва : Издательский Дом МИСиС, 2018. - 80 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-906953-14-8

5. Юрчук, С. Ю.; Методы математического моделирования Электронный ресурс : Учебное пособие / С. Ю. Юрчук. - Методы математического моделирования,2019-09-01. - Москва : Издательский Дом МИСиС, 2018. - 96 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-906953-43-8

6. Яроцкая, Е.В.; Экономико-математические методы и моделирование Электронный ресурс : учебное пособие / Е.В. Яроцкая. - Саратов : Ай Пи Эр Медиа, 2018. - 227 с. - Книга находится в базовой версии ЭБС IPRbooks. - ISBN 978-5-4486-0074-6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

по организации самостоятельной работы
по дисциплине «Моделирование и оптимизация процессов и систем сервиса»
для студентов направления подготовки 43.03.01 Сервис
Направленность (профиль) «Сервисная экономика»

Ставрополь, 2026

Методические указания по дисциплине «Моделирование и оптимизация процессов и систем сервиса» содержат задания для самостоятельной работы студентов.

Проработка предложенных вопросов позволит студентам приобрести необходимые знания в области математического моделирования.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Основными способами самостоятельной работы по изучению дисциплины «Моделирование и оптимизация процессов и систем сервиса» являются:

- чтение и конспектирование первоисточников - произведений классиков эконометрической науки;
- изучение учебников, учебно-методических пособий и другой учебной литературы;
- систематическая работа над конспектами лекций, их дополнение материалом из учебников (учебных пособий);
- подготовка докладов, научных сообщений и выступление с ними на практических занятиях;
- выполнение рефератов и эссе по темам, изучаемым в рамках дисциплины «Моделирование и оптимизация процессов и систем сервиса»;
- постановка и решение задач, выполнение заданий, рекомендованных (заданных) преподавателем;
- формулировка развернутых ответов на вопросы для подготовки к практическим занятиям;
- подготовка к зачету.

Самостоятельная работа помогает студентам:

1) приобрести знания и навыки:

- работы с научными материалами (первоисточники, дополнительная научная и учебная, специальная литература и т.д.);
- составления логических конспектов, графических изображений структуры конспектов, составление блок-схем и т.д.;
- работы со справочным материалом;
- изучения и работы с нормативными и правовыми документами, нормативно-правовыми поисковыми системами Гарант, Консультант Плюс;
- учебно-методической и научно-исследовательской работы;

- использования компьютерной техники и Интернета и др.;

2) закрепить и систематизировать знания через:

- работу с конспектом лекции;
- обработку текста, повторную работу над материалом учебника, первоисточников, дополнительной литературы;
- подготовку ответов на контрольные вопросы;
- аналитическую обработку текста;
- подготовку рефератов и эссе;
- тестирование и др.;

3) сформировать умения:

- решать ситуационные задачи и упражнения по образцу;
- готовиться к контрольным и проверочным работам, к тестированию;
- принимать участие в практических занятиях в интерактивных формах.

1. ОБЩАЯ ХАРАКТЕРИСТИКА САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Цель самостоятельной работы студента - научиться осмысленно и самостоятельно работать с учебным материалом и научной информацией, овладеть фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, сформировать основы самоорганизации и самовоспитания с тем, чтобы привить умение в дальнейшем непрерывно повышать свою профессиональную квалификацию.

Задачами СРС являются:

- систематизация и закрепление полученных теоретических знаний и практических умений бакалавров;
- углубление и расширение теоретических знаний;
- формирование умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- развитие познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развитие исследовательских умений;
- использование материала, собранного и полученного в ходе самостоятельных занятий на семинарах, на практических и лабораторных занятиях, при написании курсовых и выпускной квалификационной работ, для эффективной подготовки к итоговым зачетам.

В этой связи важнейшая задача учебного процесса - научить студентов мыслить и усваивать знания. Учащимся необходимо превратиться из пассивных потребителей знаний в активных их творцов, умеющих грамотно сформулировать проблему, проанализировать пути ее решения, найти оптимальный результат и доказать его правильность.

Все ранее перечисленное предполагает ориентацию на активные методы овладения знаниями, развитие творческих способностей студентов, переход от поточного обучения к индивидуализированному, с учетом потребностей и возможностей личности. Поэтому формы учебного процесса и все методики обучения совершенствуются с целью активизации самостоятельной работы студентов (СРС).

Под самостоятельной работой студентов понимается совокупность всей самостоятельной деятельности студентов как в учебной аудитории, так и за ее пределами, в контакте с преподавателем и в его отсутствие.

Направления реализации самостоятельной работы:

1. В рамках аудиторных занятий - на лекциях, практических занятиях и в ходе выполнения контрольно-самостоятельных работ.
2. Через контакт с преподавателем вне расписания - консультации по учебным вопросам, творческие контакты, ликвидации задолженностей, отчет студента о ходе выполнения учебных и творческих задач.
3. В рамках работы с электронными библиотечными системами (ЭБС типа znanium.com или e.lanbook.com).

Самостоятельная работа по дисциплине «Моделирование и оптимизация процессов и систем сервиса» выполняется с целью получения и закрепления знаний, приобретенных при изучении теоретического материала.

2. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ИЗУЧЕНИЮ ТЕОРЕТИЧЕСКОГО МАТЕРИАЛА, СОБЕСЕДОВАНИЕ, КОМПЛЕКТ ЗАДАЧ

Изучение любого раздела или темы следует начинать с ознакомления с вопросами плана изучения темы. Теоретический материал представляет собой конспект лекций, содержащий необходимый набор утверждений и формул (без детальных подробностей), но с подробным обоснованием их использования при решении конкретных экономических задач. При изучении материала необходимо помимо лекционных материалов использовать рекомендуемую основную и дополнительную литературу для лучшего усвоения материала.

Осваивать теорию следует в соответствии с той последовательностью, которая представлена в плане лекции. Методика работы с литературой предусматривает ведение записи прочитанного в виде плана - конспекта, опорного конспекта. Это позволит сделать знания системными, зафиксировать и закрепить их в памяти.

Для успешного освоения дисциплины, необходимо самостоятельно детально изучить представленные темы по рекомендуемым источникам информации:

Этапы собеседования

Собеседование можно условно подразделить на два этапа:

Подготовительный этап, включающий подготовку студента к собеседованию путем изучения предмета исследования и выбор вопроса (проблемы);

Основной этап, включающий изложение результатов изученного материала в виде связного текста и ответы на встречные вопросы преподавателя.

Подготовка студента к собеседованию начинается с выбора вопроса (проблемы) для изучения. Вопрос (проблема) для собеседования распределяется преподавателем или выбирается студентами из предлагаемой примерной тематики, которая соответствует рабочей программе дисциплины. Вместе с тем студенту предоставляется право взять иной вопрос, который является начальным этапом или продолжением его практической научно-исследовательской работы, учитывает его личные интересы и склонности, способности, а также уровень теоретических знаний и профессиональной практической подготовки и дает возможность творчески подойти к его разработке.

Поиск источников. Грамотно сформулированной темой фиксируется предмет изучения; задача студента - найти информацию, относящуюся к данному предмету и разрешить поставленную проблему. Выполнение этой задачи начинается с поиска источников. На этом этапе необходимо вспомнить, как работать с энциклопедиями и энциклопедическими словарями (обращать особое внимание на список литературы, приведенный в конце тематической статьи); как работать с систематическими и алфавитными каталогами библиотек; как оформлять список литературы (выписывая выходные данные книги и отмечая библиотечный шифр).

Работа с источниками. Работу с источниками надо начинать с ознакомительного чтения, т.е. просмотреть текст, выделяя его структурные единицы. При ознакомительном чтении закладками отмечаются те страницы, которые требуют более внимательного изучения.

В зависимости от результатов ознакомительного чтения выбирается дальнейший способ работы с источником. Если для разрешения поставленной задачи требуется изучение некоторых фрагментов текста, то используется метод выборочного чтения. Если источник не содержит подробного оглавления, следует обратить внимание на предметные и именные указатели.

Избранные фрагменты или весь текст (если он целиком имеет отношение к теме) требуют вдумчивого, неторопливого чтения с «мысленной проработкой» материала. Такое чтение предполагает выделение: главного в тексте; основных аргументов; выводов. Особое внимание следует обратить на то, вытекает тезис из аргументов или нет.

Необходимо также проанализировать, какие из утверждений автора носят проблематичный, гипотетический характер и уловить скрытые вопросы.

Понятно, что умение таким образом работать с текстом приходит далеко не сразу. Наилучший способ научиться выделять главное в тексте, улавливать проблематичный характер утверждений, давать оценку авторской позиции – это сравнительное чтение, в ходе которого студент знакомится с различными мнениями по одному и тому же вопросу, сравнивает весомость и доказательность аргументов сторон и делает вывод о наибольшей убедительности той или иной позиции.

Создание вспомогательных конспектов к собеседованию. Подготовительный этап работы может завершаться созданием конспектов, фиксирующих основные тезисы и аргументы, статистическую или ссылочную информацию. Написание вспомогательных конспектов не является обязанностью студента при подготовке к собеседованию, и никак не регулируется оформление таких конспектов.

Основной этап. Студент должен излагать известный материал грамотно, логично, с использованием научных терминов и научного стиля изложения. Студенту не следует отклонятся от установленного предмета изучения и отходить от круга изученных им вопросов входе подготовки к собеседованию. Преподавателем могут задаваться вопросы по существу излагаемого материала, а студент должен своевременно, полно и правильно на них отвечать.

Критерии оценки знаний

Дата, время и место собеседования назначаются преподавателем.

Оценка знаний студентов по конкретному вопросу в ходе собеседования производится следующим образом:

Оценка «отлично» выставляется студенту, если материал по существу вопроса (проблемы) изложен верно, выводы по результатам самостоятельного изучения студентом сделаны верные, студент использовал требуемый стиль изложения, ответил на все вопросы преподавателя верно.

Оценка «хорошо» выставляется студенту, если материал по существу вопроса (проблемы) изложен верно и/или не полностью, выводы по результатам самостоятельного изучения студентом сделаны верные, студент не использовал требуемый стиль изложения, ответил не на все вопросы преподавателя верно.

Оценка «удовлетворительно» выставляется студенту, если материал по существу вопроса (проблемы) изложен с ошибками и не полностью, выводы по результатам самостоятельного изучения студентом сделаны неверные или не сделаны, студент не использовал требуемый стиль изложения, не ответил верно на большую часть вопросов преподавателя.

Оценка «неудовлетворительно» выставляется студенту, если материал по существу вопроса (проблемы) изложен неверно либо не изложен, студент не использовал требуемый стиль изложения, не ответил верно на большую часть вопросов преподавателя.

Тема: Базовые понятия эконометрики.
Вопросы:
Индексы и их использование в экономико-статистических исследованиях. Экономические модели и статистические методы Основные типы эконометрических моделей. Стохастическая природа экономических данных. Методы сбора статистических данных, подготовка их к использованию в моделях.

Тема: Парный регрессионный анализ.
Вопросы:
Общее понятие об индексах и значение индексного метода анализа. Индексы количественных и качественных показателей. Индивидуальные и общие индексы. Индексы в микроэкономическом и макроэкономическом анализе. Агрегатная форма общего рынка. Взаимосвязи индексов товарооборота. Территориальные индексы. Примеры.

Тема: Множественный регрессионный анализ.
Вопросы:
Элементы корреляционного анализа. Множественная корреляция. Корреляционные и регрессионные уравнения.

Тема: Регрессионные модели с переменной структурой.
Вопросы:
Вероятностные соотношения; статистическая зависимость (независимость) случайных переменных.
Анализ линейной статистической связи экономических данных, корреляция; вычисление коэффициентов корреляции; связь признаков, измеренных в разных шкалах.

Тема: Специфика построения динамических регрессионных моделей.
Вопросы:
Нелинейная статистическая вязь, корреляционное отношение. Корреляционные и регрессионные уравнения. Непараметрические методы оценки корреляционной связи.

Тема: Гетероскедастичность в регрессионных моделях.
Вопросы:
Многофакторный дисперсионный анализ.
Построение многофакторных экономических моделей. Однофакторный и многофакторный дисперсионный анализ; основные положения и задачи дисперсионного анализа.
Сравнение нескольких средних; общая, факторная и остаточная суммы квадратов и соответствующие дисперсии. Связь однофакторного и двухфакторного анализа.

Тема: Эконометрические модели в виде систем одновременных уравнений.
Вопросы:
Параметрические методы проверки гипотез.
Непараметрические методы проверки гипотез.

Тема: Моделирование одномерных временных рядов.
Вопросы:
Неодинаковое число наблюдений, параметрические и непараметрические методы проверки гипотез об отсутствии эффектов обработки данных. Автокорреляция остатков при моделировании временных рядов

КОМПЛЕКТ ЗАДАЧ

1 Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Валовой доход за год, млн. руб.	Среднегодовая стоимость, млн. руб. основных фондов	Среднегодовая стоимость, млн. руб. оборотных средств
1	203	118	105
2	63	28	56
3	45	17	54
4	113	50	63
5	121	56	28
6	88	102	50
7	110	116	54
8	56	124	42
9	80	114	36
10	237	154	106

2. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Оборот капитала, млрд.долларов США	Использованный капитал, млрд. долларов США
1	6,6	6,9	83,6
2	3,0	18,0	6,5
3	6,5	107,9	50,4
4	3,3	16,7	15,4
5	0,1	79,6	29,6
6	3,6	16,2	13,3
7	1,5	5,9	5,9
8	5,5	53,1	27,1
9	2,4	18,8	11,2
10	3,0	35,3	16,4

4. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Валовой доход за год, млн. руб.	Среднегодовая стоимость, млн. руб. основных фондов	Среднегодовая стоимость, млн. руб. оборотных средств
1	237	154	106
2	160	115	88
3	75	98	46

4	203	118	105
5	63	28	56
6	45	17	54
7	113	50	63
8	121	56	28
9	88	102	50
10	110	116	54

5. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Оборот капитала, млрд.долларов США	Использованный капитал, млрд. долларов США
1	0,8	6,8	3,2
2	1,8	27,0	13,0
3	0,9	12,4	6,9
4	1,1	17,7	15,0
5	1,9	12,7	11,9
6	0,9	21,4	1,6
7	1,3	13,5	8,6
8	2,0	13,4	11,5
9	0,6	4,2	1,9
10	0,7	15,5	5,8

6. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Численность служащих, тыс. чел.	Рыночная капитализация компании, млрд. долларов США
1	0,9	43,0	40,9
2	1,7	64,7	40,5
3	0,7	24,0	38,9
4	1,7	50,2	38,5
5	2,6	106,0	37,3
6	1,3	96,6	26,5
7	4,1	347,0	37,0
8	1,6	85,6	36,8
9	6,9	745,0	36,3
10	0,4	4,1	35,3

7. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Численность служащих, тыс. чел.	Рыночная капитализация компании, млрд. долларов США
1	1,3	26,8	35,3
2	1,9	42,7	35,0

3	1,9	61,8	26,2
4	1,4	212,0	33,1
5	0,4	105,0	32,7
6	0,8	33,5	32,1
7	1,8	142,0	30,5
8	0,9	96,0	29,8
9	1,1	140,0	25,4
10	1,9	59,3	29,3

8. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Численность служащих, тыс. чел.	Рыночная капитализация компании, млрд. долларов США
1	0,8	33,5	32,1
2	1,8	142,0	30,5
3	0,9	96,0	29,8
4	1,1	140,0	25,4
5	1,9	59,3	29,3
6	0,9	131,0	29,2
7	1,3	70,7	29,2
8	2,0	65,4	29,1
9	0,6	23,1	27,9
10	0,7	80,8	27,2

9. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Оборот капитала, млрд.долларов США	Использованный капитал, млрд. долларов США
1	6,6	6,9	83,6
2	3,0	18,0	6,5
3	6,5	107,9	50,4
4	3,3	16,7	15,4
5	0,1	79,6	29,6
6	3,6	16,2	13,3
7	1,5	5,9	5,9
8	5,5	53,1	27,1
9	2,4	18,8	11,2
10	3,0	35,3	16,4

10. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер /название страны	Индекс человеческого развития	ВВП 2007 г., % к 2000 г.	Расходы на конечное потребление в текущих ценах, % к ВВП

1(Австрия)	0,904	115,0	75,5
2(Австралия)	0,922	123,0	78,5
3(Белоруссия)	0,763	74,0	78,4
4(Бельгия)	0,923	111,0	77,7
5(Великобритания)	0,918	113,0	84,4
6(Германия)	0,906	110,0	75,9
7(Дания)	0,905	110,0	75,9
8(Индия)	0,545	146,0	67,5
9(Испания)	0,894	113,0	78,2
10(Италия)	0,900	108,0	78,

11. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии.

Номер /название страны	Индекс человеческого развития	Расход домашнего хозяйства, % к ВВП	Валовое накопление, % к ВВП
1(Австрия)	0,904	56,1	25,2
2(Австралия)	0,922	61,8	21,8
3(Белоруссия)	0,763	59,1	25,7
4(Бельгия)	0,923	63,3	17,8
5(Великобритания)	0,918	64,1	15,9
6(Германия)	0,906	57,0	22,4
7(Дания)	0,905	50,7	20,6
8(Индия)	0,545	57,1	25,2
9(Испания)	0,894	62,0	20,7
10(Италия)	0,900	61,8	17,5

12. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Валовой доход за год, млн. руб.	Среднегодовая стоимость, млн. руб. основных фондов	Среднегодовая стоимость, млн. руб. оборотных средств
1	203	118	105
2	63	28	56
3	45	17	54
4	113	50	63
5	121	56	28
6	88	102	50
7	110	116	54
8	56	124	42
9	80	114	36
10	237	154	106

13. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Оборот капитала, млрд.долларов США	Использованный капитал, млрд. долларов США
1	6,6	6,9	83,6
2	3,0	18,0	6,5
3	6,5	107,9	50,4
4	3,3	16,7	15,4

5	0,1	79,6	29,6
6	3,6	16,2	13,3
7	1,5	5,9	5,9
8	5,5	53,1	27,1
9	2,4	18,8	11,2
10	3,0	35,3	16,4

14. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Валовой доход за год, млн. руб.	Среднегодовая стоимость, млн. руб.основных фондов	Среднегодовая стоимость, млн. руб. оборотных средств
1	237	154	106
2	160	115	88
3	75	98	46
4	203	118	105
5	63	28	56
6	45	17	54
7	113	50	63
8	121	56	28
9	88	102	50
10	110	116	54

15. Построить с помощью MS Excel модель множественной корреляции используя уравнение множественной регрессии. Найти коэффициент парной корреляции. Найти коэффициент множественной корреляции.

Номер предприятия	Чистый доход, млрд.долларов США	Оборот капитала, млрд.долларов США	Использованный капитал, млрд. долларов США
1	0,8	6,8	3,2
2	1,8	27,0	13,0
3	0,9	12,4	6,9
4	1,1	17,7	15,0
5	1,9	12,7	11,9
6	0,9	21,4	1,6
7	1,3	13,5	8,6
8	2,0	13,4	11,5
9	0,6	4,2	1,9
10	0,7	15,5	5,8

3. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПОДГОТОВКЕ ДОКЛАДА

К самостоятельной работе относится написание и защита доклада в семестре. Подготовка доклада по дисциплине – один из основных этапов учебного процесса в обучении студентов, которым необходимо приобрести навыки самостоятельного исследования и представления его результатов. Тема выбирается студентом самостоятельно по согласованию с преподавателем.

Примерные темы доклада:

1. Объект и предмет эконометрики. Методы исследования
2. Оценка параметров известной теоретической модели.
3. Выявление эмпирически неизвестных закономерностей и проверка статистических гипотез.
4. Методы статистического анализа в эконометрике.
5. Экономические модели и статистические методы.
6. Основные типы эконометрических моделей.
7. Методы сбора статистических данных и подготовка их к использованию в моделях.
8. Способы представления экономических данных.
9. Общее понятие об индексах и значение индексного метода анализа.
10. Индексы количественных и качественных показателей.
11. Индивидуальные и общие индексы.
12. Индексы в микроэкономическом и макроэкономическом анализе.
13. Агрегатная форма общего рынка.
14. Взаимосвязи индексов товарооборота.
15. Территориальные индексы.
16. Содержание корреляционного анализа.
17. Вероятностные соотношения; статистическая зависимость (независимость) случайных переменных.
18. Анализ линейной статистической связи экономических данных.
19. Корреляция; вычисление коэффициентов корреляции.
20. Связь признаков, измеренных в разных шкалах.
21. Построение многофакторных экономических моделей.
22. Однофакторный и многофакторный дисперсионный анализ.
23. Основные положения и задачи дисперсионного анализа.
24. Сравнение нескольких средних.
25. Связь однофакторного и двухфакторного анализа.
26. Таблицы дисперсионного анализа.

27. Параметрические и непараметрические методы проверки гипотез об отсутствии эффектов обработки данных.

28. Проблемы оценивания связи экономических переменных и регрессионнный анализ.

29. Линейная регрессия; статистический анализ модели; прогнозирование.

30. Нелинейная регрессия. Типовые регрессионные модели в эконометрике.

31. Эконометрический регрессионный анализ макроэкономических моделей.

32. Оценка функции потребления.

33. Производственная функция Кобба-Дугласа.

34. Макромодель Клейна.

35. Эконометрическая модель фирмы.

36. Случайные процессы.

37. Временные ряды.

38. Статистические показатели динамики социально-экономических явлений.

39. Анализ временных рядов.

40. Изучение динамики в экономической деятельности.

41. Стационарные и нестационарные временные ряды.

42. Детерминированная и случайная составляющие временного ряда.

43. Числовые характеристики временных рядов.

44. Практический анализ временных рядов.

45. Средние показатели в рядах динамики. метод выделения тренда.

46. Линейные модели временных рядов; процессы скользящего среднего.

47. Изучение сезонных колебаний. Экстраполяция и прогнозирование.

48. Структура стационарного временного ряда; автокорреляционная функция.

49. Полный и дробный факторный эксперимент.

50. Планирование в условиях неоднородности, анализ качественных факторов.

51. Последовательный анализ.

52. Многомерный статистический анализ.

53. Факторный анализ.

Общие рекомендации по подготовке доклада

Доклад должен включать в себя введение, основную часть и заключение.

Во введении необходимо отразить обоснование актуальности выбранной темы, краткое описание текущего состояния проблемы. В нем студент должен указать цель и задачи работы, объект исследования, элементы новизны, введенные в процессе написания работы. Необходимо перечислить проблемы, которые должны быть решены в рамках выбранной темы.

Основная часть доклада должна содержать вопросы, предусмотренные в плане работы. В ней необходимо отразить теоретические основы, раскрывающие суть проблемы, проанализировать собранные материалы, характеризующие практическую сторону объекта исследования. Этот раздел может содержать рабочие таблицы, диаграммы (диаграммы и другие материалы.

В заключение необходимо отразить выводы и предложения, полученные в результате предыдущей работы. Они должны быть сформулированы четко и точки.

Список литературы включает в алфавитном порядке список современных законов и нормативных актов, соответствующей научной литературы, научных работ, статистических сборников и других источников, выпущенных не ранее пяти лет.

Оформление доклада и порядок защиты

Доклад должен иметь титульный лист, план работы, непосредственно текст доклада, список литературы и приложения.

Объем работы - 20-25 страниц пронумерованы компьютерного текста, шрифт, 14, интервал 1,5, поля 2-3 см приложений имеют внутренний (частный) нумерацию страниц. Иллюстрации, фотографии, рисунки, графики, фотографии, которые появляются на тексте, должны быть пронумерованы. Выполненный доклад проверяется преподавателем. Если доклад оформлен согласно предъявляемым требованиям, то работа допускается к защите, о чем преподавателем делаются записи на титульном листе работы. Если доклад имеет отрицательный отзыв, то документ возвращается на доработку с последующим представлением о его повторном рассмотрении.

Критерии оценки доклада

Оценка «отлично» ставится, если выполнены все требования к написанию и защите доклада: обозначена проблема и обоснована её актуальность, сделан краткий анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция, сформулированы выводы, тема раскрыта полностью, выдержан объём, соблюдены требования к внешнему оформлению, даны правильные ответы на дополнительные вопросы.

Оценка «хорошо» - основные требования к докладу и его защите выполнены, но при этом допущены недочёты. В частности, имеются неточности в изложении материала; отсутствует логическая последовательность в суждениях; не выдержан объём эссе; имеются упущения в оформлении; на дополнительные вопросы при защите даны неполные ответы.

Оценка «удовлетворительно» - имеются существенные отступления от требований к реферированию.

Оценка «неудовлетворительно» - Тема доклада не раскрыта, обнаруживается существенное непонимание проблемы, доклад выпускником не представлен.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Список рекомендованной литературы

Перечень основной литературы:

1. Афонин, В. В. Моделирование систем : учебное пособие / В. В. Афонин, С. А. Федосин. - 4-е изд. - Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2024. - 269 с. - ISBN 978-5-4497-2413-7. - Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. - URL: https://www.iprbookshop.ru/133951.html. Книга находится в премиум-версии ЭБС IPR BOOKS.

2. Яроцкая, Е. В. Экономико-математические методы и моделирование : учебное пособие / Е. В. Яроцкая. - 2-е изд. - Москва : Ай Пи Ар Медиа, 2026 . - 196 с. - ISBN 978-5-4497-3855-4. - Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. - URL: https://www.iprbookshop.ru/145188.html. Книга находится в премиум-версии ЭБС IPR BOOKS.

Перечень дополнительной литературы:

1. Наумов, И.В.; Эконометрика. Экономическое моделирование социально-экономических процессов в территориальных системах: учебное пособие / И. В. Наумов, Н. Л. Никулина. - Эконометрика. Экономическое моделирование социально-экономических процессов в территориальных системах, Весь срок охраны авторского права. - Электрон. дан. (1 файл). - Москва : Ай Пи Ар Медиа, 2021. - 127 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4497-1408-4,

2. Орлов, А. И. Эконометрика : учебное пособие / А. И. Орлов. - Эконометрика,2021-12-05. - Электрон. дан. (1 файл). - Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. - 676 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4497-0362-0,

3. Лихтенштейн, В.Е.; Математическое моделирование экономической справедливости Электронный ресурс : монография / Г.В. Росс / В.Е. Лихтенштейн. - Саратов : Ай Пи Эр Медиа, 2018. - 129 с. - Книга находится в базовой версии ЭБС IPRbooks. - ISBN 978-5-4486-0350-1

4. Петров, А. Е.; Математические модели принятия решений Электронный ресурс : Учебно-методическое пособие / А. Е. Петров. - модели принятия решений,2019-09-01. - Москва : Издательский Дом МИСиС, 2018. - 80 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-906953-14-8

5. Юрчук, С. Ю.; Методы математического моделирования Электронный ресурс : Учебное пособие / С. Ю. Юрчук. - Методы математического моделирования,2019-09-01. - Москва : Издательский Дом МИСиС, 2018. - 96 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-906953-43-8

6. Яроцкая, Е.В.; Экономико-математические методы и моделирование Электронный ресурс : учебное пособие / Е.В. Яроцкая. - Саратов : Ай Пи Эр Медиа, 2018. - 227 с. - Книга находится в базовой версии ЭБС IPRbooks. - ISBN 978-5-4486-0074-6