

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
ДОНЕЦКОЙ НАРОДНОЙ РЕСПУБЛИКИ  
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
«ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**Л.Д. СЛЕПНЕВА, Е.А. РЫТИКОВА**

## **ЭКОНОМЕТРИКА**

Учебное пособие  
для обучающихся образовательных учреждений  
высшего профессионального образования  
по направлению подготовки 38.03.01 «Экономика»

Донецк

2019

УДК 330.43 (075.8)  
ББК 65в6я73  
С47

Рекомендовано Ученым советом  
ГОУВПО «Донецкий национальный технический университет»  
в качестве учебного пособия для обучающихся образовательных учреждений  
высшего профессионального образования  
по направлению подготовки 38.03.01 «Экономика»  
(Протокол №3 от 26.04.2019 г.)

**Рецензенты:**

Волощенко Л.М. – доктор экономических наук, профессор, доцент, заведующая кафедрой финансов ГОУ ВПО «Донецкая академия управления и государственной службы при Главе Донецкой Народной Республики;

Кучер В.А. – доктор экономических наук, профессор, профессор кафедры экономики и маркетинга ГОУ ВПО «Донецкий национальный технический университет»;

Рыбникова Г.И. – кандидат экономических наук, доцент, доцент кафедры «Экономическая теория и государственное управление» ГОУ ВПО «Донецкий национальный технический университет»

**Авторы:**

Слепнева Л.Д. – кандидат экономических наук, доцент, доцент кафедры «Финансы и экономическая безопасность» ГОУ ВПО «ДОННТУ»;

Рытикова Е.А. – старший преподаватель кафедры «Финансы и экономическая безопасность» ГОУ ВПО «ДОННТУ»

**Слепнева, Л. Д.**

С47 Эконометрика : учебное пособие / Л.Д. Слепнева, Е.А. Рытикова; – ГОУВПО «ДОННТУ». Донецк : ГОУВПО «ДОННТУ». 2019. – 139 с. : ил., табл.  
ISBN

В учебном пособии рассмотрены теоретические основы дисциплины «Эконометрика» в объёме, рекомендуемом Государственным образовательным стандартом высшего профессионального образования и Основной образовательной программой высшего профессионального образования по направлению подготовки 38.03.01 «Экономика». Основное внимание уделено изложению особенностей применения эконометрических методов и современных программных средств для анализа экономических процессов.

Пособие предназначено для обучающихся очной и заочной формы обучения.

УДК 330.43 (075.8)  
ББК 65в6я73

ISBN

© Слепнева Л.Д., Рытикова Е.А., 2019

© ГОУВПО «Донецкий национальный технический университет»

## ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	5
ТЕМА 1. ОСНОВНЫЕ АСПЕКТЫ ЭКОНОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ	7
1.1 Место эконометрики среди смежных дисциплин	7
1.2 Понятие регрессии	11
1.3 Понятие корреляции	14
1.4 Понятие эконометрической модели и основные задачи эконометрического моделирования	17
Вопросы для самоконтроля	23
ТЕМА 2. МОДЕЛИ ПАРНОЙ РЕГРЕССИИ	24
2.1 Построение уравнения парной регрессии методом наименьших квадратов	24
2.2 Оценка тесноты связи и полноты модели	29
2.3 Проверка статистической значимости (надежности) результатов оценивания	31
2.4 Точечный и интервальный прогноз по уравнению линейной регрессии	34
2.5 Коэффициент эластичности	35
2.6 Выбор наилучшей функции регрессии	36
2.7 Построение эконометрической модели с использованием EXCEL Вопросы для самоконтроля	39 42
ТЕМА 3. МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ	43
3.1 Оценивание параметров многофакторной регрессии методом наименьших квадратов	43
3.2 Использование модуля <i>Анализ данных</i> EXCEL	51
3.3 Проверка качества регрессионной модели (коэффициент детерминации), информационные критерии	58
Вопросы для самоконтроля	62
ТЕМА 4. МУЛЬТИКОЛЛИНЕАРНОСТЬ	63
4.1 Понятие мультиколлинеарности	63
4.2 Исследование мультиколлинеарности	65
4.3 Устранение мультиколлинеарности	66
Вопросы для самоконтроля	71

ТЕМА 5. ИСПОЛЬЗОВАНИЕ ФИКТИВНЫХ ПЕРЕМЕННЫХ В ЭКОНОМЕТРИЧЕСКОМ АНАЛИЗЕ	72
5.1 Введение фиктивных переменных в линейную модель регрессии	72
5.2 Проверка регрессионной однородности двух групп наблюдений	78
Вопросы для самоконтроля	80
ТЕМА 6. РЕГРЕССИОННЫЕ МОДЕЛИ С ГЕТЕРОСКЕДАСТИЧНЫМИ ОСТАТКАМИ. ТЕСТИРОВАНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ	81
6.1 Понятие гетероскедастичности	81
6.2 Тестирование гетероскедастичности	83
Вопросы для самоконтроля	92
ТЕМА 7. ЛИНЕЙНЫЕ МОДЕЛИ С АВТОКОРРЕЛИРОВАННЫМИ ОСТАТКАМИ. ТЕСТИРОВАНИЕ АВТОКОРРЕЛЯЦИИ	93
7.1 Понятие автокорреляции	93
7.2 Тестирование автокорреляции	94
Вопросы для самоконтроля	101
ТЕМА 8. ОБОБЩЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ. МЕТОД ЭЙТКЕНА	102
8.1 Понятие обобщенной линейной регрессионной модели	102
8.2 Обобщенный метод наименьших квадратов (ОМНК)	105
Вопросы для самоконтроля	115
ТЕМА 9. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ (СОУ)	116
9.1 Необходимость использования СОУ	116
9.2 Переменные в системах одновременных уравнений	117
9.3 Виды систем уравнений	120
9.4 Структурная и приведенная формы модели	122
9.5 Проблема идентифицируемости. Условия идентифицируемости уравнений системы	125
9.6 Оценивание параметров структурной формы эконометрической модели	129
Вопросы для самоконтроля	132
ЛИТЕРАТУРА	133
ПРИЛОЖЕНИЕ А. Математико-статистические таблицы	135
Предметный указатель	139

## ПРЕДИСЛОВИЕ

В основу учебного пособия положен опыт преподавания дисциплины «Эконометрика» в Донецком национальном техническом университете для обучающихся по направлению подготовки 38.03.01 «Экономика» очной и заочной форм обучения.

Учебное пособие разработано в соответствии с требованиями Государственного образовательного стандарта высшего профессионального образования и Основной образовательной программы высшего профессионального образования по направлению подготовки 38.03.01 «Экономика» и содержит систематическое изложение базовых понятий эконометрики, формулировку типовых задач и обоснование методических подходов к их решению с использованием ЭВМ.

Дисциплина «Эконометрика» призвана научить различным способам выражения связей и закономерностей через эконометрические модели, основанные на данных статистических наблюдений. Эконометрический подход предусматривает анализ соответствия выбранной модели изучаемому объекту, рассмотрение причин, приводящих к необходимости пересмотра моделей на основе более точной системы представлений.

Цель дисциплины – формирование системы знаний по методологии и инструментарию построения и использования разных типов эконометрических моделей.

Задачи – изучить основные принципы и инструментарий постановки задач, построения эконометрических моделей, методов их оценивания и анализа с целью использования в экономике.

Дисциплина «Эконометрика» опирается на курсы «Макроэкономика», «Микроэкономика», «Математический анализ», «Линейная алгебра», «Теория вероятностей и математическая статистика».

В свою очередь, «Эконометрика» выступает в качестве базы для изучения дисциплин «Менеджмент», «Маркетинг», «Финансовый менеджмент», «Финансовый анализ», при прохождении учебной и производственной практики, а также при написании бакалаврской дипломной работы.

Учебное пособие включает 9 глав, в которых изложены теоретические предпосылки эконометрического исследования.

Основные аспекты эконометрического моделирования – предмет рассмотрения первой главы.

Вторая и третья главы посвящены вопросам построения и анализа классических моделей парной и множественной линейной регрессии. Здесь же показаны способы линеаризации модели, частная корреляция, построение прогнозов.

В главах 4 и 5 рассматриваются проблемы, связанные с использованием регрессионных моделей, в частности, мультиколлинеарность (влияние на результаты эконометрического моделирования, тестирование, пошаговая регрессия как один из способов смягчения или устранения мультиколлинеарности), а также применение фиктивных переменных для характеристики влияния качественных признаков.

В главах 6 и 7 представлены обобщенные регрессионные модели. Исследуются вопросы, возникающие из-за нарушения основных предпосылок классической регрессии – гетероскедастичности и автокоррелированности остатков регрессии.

Глава 8 посвящена рассмотрению обобщенного метода наименьших квадратов и применению его для получения эффективных оценок параметров модели при гетероскедастичности или автокоррелированности остатков.

В главе 9 излагаются вопросы построения и использования эконометрических моделей, представленных системой одновременных уравнений.

## ТЕМА 1. ОСНОВНЫЕ АСПЕКТЫ ЭКОНОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ

### 1.1 Место эконометрики среди смежных дисциплин

Поскольку эксперимент в экономике невозможен, то основным средством ее изучения является экономико-математическая модель. Составными элементами такой модели оказываются отвлеченные экономические понятия, такие переменные, как цена, количество товаров и доходы, которые связаны между собой математическими соотношениями: алгебраическими уравнениями или кривыми на графиках, причем связаны не произвольно, а строго определенным образом.

Любая модель является лишь некоторым подобием реального явления или процесса, отражая его основные сущностные характеристики, но оставляя без внимания множество частных фактов, имеющих место в реальности. Американский экономист и статистик Э. Кейн отмечал: «С помощью этих моделей можно понять много, но никогда не удастся понять все. Это имеет свои хорошие и плохие стороны. С одной стороны, никому не требуется (да и не хочется) знать о чем-либо все. Как правило, нужно знать только то, что действительно важно. С другой стороны, никогда нельзя быть уверенным в том, что в некой модели удалось учесть все те факторы, которые особенно важны при теоретическом исследовании проблемы (или для прикладных задач). Здесь-то и начинается эконометрия! Сопоставляя теоретические экономические модели с реальными данными, эконометрия может ответить на важнейший вопрос о соответствии между ними и помочь создать целую систему не очень сложных и вполне адекватных экономических моделей» [8, с. 16].

29 декабря 1930 года в Кливленде было основано «Эконометрическое общество», которое определило себя как «международное общество для развития экономической теории в ее связи со статистикой и математикой». Основателями этого общества являются Рагнар Фриш, лауреат Нобелевской премии в области экономики, и Чарльз Ф. Рус.

В журнале «Эконометрика», основанном в 1933 г. Р. Фришем (1895-1973), дано следующее определение эконометрики: «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 трёх отправных точек – статистика, экономическая теория и математика – необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. Это единство всех трёх составляющих. И это единство образует эконометрику» [1].

В современном понимании **эконометрика – научная дисциплина, разрабатывающая и использующая методы, модели, позволяющие придавать конкретное количественное выражение общим (качественным) закономерностям экономической теории на базе экономической статистики с использованием математико-статистического инструментария** [2, с. 597].

Из определения эконометрии следует, что ее происхождение и главное назначение – это экономические и социально-экономические приложения, а именно: модельное описание *конкретных количественных взаимосвязей*, существующих между анализируемыми показателями.

Следующая схема (рис. 1.1) позволит лучше разобраться в сущности рассматриваемого вопроса.



Рисунок 1.1 - Эконометрия и место ее в ряду смежных дисциплин [2]

Экономическая теория (в рамках эконометрии) интересует исследователя не столько выявлением объективно существующих (на качественном уровне) экономических законов и связей между экономическими показателями, сколько подходами к их формализации, включающими в себя *методы спецификации соответствующих моделей с учетом проблемы их идентифицируемости*.

При рассмотрении экономической статистики как составной части эконометрии обычно имеют в виду тот аспект этой самостоятельной дисциплины, который непосредственно связан с *информационным обеспечением анализируемой эконометрической модели*.

Под математико-статистическим инструментарием эконометрии подразумеваются отдельные разделы математической статистики: *классическая и обобщенная линейные модели регрессионного анализа, анализ временных рядов, построение и анализ системы одновременных уравнений*.

*Область исследований эконометрии* определяется моделями математической экономики. Известно, что математическая экономика переводит экономические проблемы на язык математики для более удобного обращения со сложной системой понятий политической экономии и производных от нее или подчиненных ей дисциплин. Эконометрия отыскивает числовые значения параметров моделей математической экономики.

*Методом эконометрии* является метод, предложенный математической статистикой.

*Объектом эконометрии* является экономика, но в особом, количественном аспекте.

*Предмет эконометрии* - новые методы получения статистических выводов в рамках математических моделей взаимосвязи экономических показателей.

**Целью эконометрии** является открытие механизмов преобразования некоторых величин (X) в другие (Y) путем формирования гипотез об этих механизмах в виде математических моделей, приспособления к ним числовых данных измерений и выбора в качестве «истинной» той модели, которая обнаруживает максимальное правдоподобие.

В процессе проведения эконометрического исследования решаются следующие **основные задачи** (рис. 1.2):

Спецификация	Идентификация (параметризация)	Верификация	Использование
<p>Построение эконометрических моделей, т.е. представление экономических явлений в математической форме, удобной для проведения эмпирического анализа</p>	<p>Оценка параметров построенной модели по выборочным данным, делающих выбранную модель наиболее адекватной реальным данным</p>	<p>Оценка качества модели (найденных оценок параметров и самой модели в целом)</p>	<p>Применение построенных моделей для объяснения поведения исследуемых экономических показателей, а также для прогнозирования и осмысленного проведения экономической политики</p>

Рисунок 1.2 - Основные задачи эконометрического исследования

Решение перечисленных задач (рис. 1.2) составляет содержание эконометрического исследования. Наиболее существенной задачей считается верификация – оценка качества эконометрической модели, которая включает проверку статистической значимости результатов оценивания и наличия требуемых свойств у оценок параметров.

Для того чтобы разобраться в сущности эконометрического моделирования, необходимо более детально рассмотреть вопросы, определяющие использование аппарата математико-статистического инструментария, в частности, регрессионного и корреляционного анализа.

## 1.2 Понятие регрессии

Большинство экономических явлений и процессов представляют собой результат многих одновременно и совокупно действующих причин, среди которых могут находиться главные и второстепенные. Так, на величину себестоимости единицы продукции влияют объем производства, применяемая технология, уровень производительности труда. Производительность труда, которая выступает как причинный фактор в формировании величины себестоимости, в свою очередь – следствие различных причин, таких как уровень развития техники и технологии, использования оборудования, научной организации труда и т.д. Кроме того, причинному действию и определяемому им следствию в той или иной степени присуща случайность.

Поскольку случайности и второстепенные факторы не могут быть исключены из опытных (эмпирических) данных, зависимость приобретает *стохастический характер*, за которым может быть скрыта функциональная (однозначная) связь. Поэтому исследование экономического процесса обычно начинают с установления существенных в данных условиях причин и выражения основных причинных отношений в количественной форме. Эта задача решается с помощью **регрессионного анализа**, который занимает центральное место среди методов математико-статистического инструментария эконометрики.

***Под регрессией понимают одностороннюю стохастическую зависимость одной случайно переменной от другой или нескольких переменных***

При стохастической (или вероятностной) зависимости, в отличие от функциональной, каждому фиксированному значению аргумента соответствует определенное статистическое распределение значений функции. Это обусловлено тем, что зависимая переменная подвержена влиянию ряда неучтенных факторов, а также наличием случайных ошибок, возникающих при измерении переменных. Значения зависимой переменной являются, таким образом, реализациями случайной величины.

Односторонняя стохастическая зависимость проявляется в том, что функция регрессии не обладает свойством обратимости, как функциональная зависимость. Например, функция  $X = 1/4 * Y$  является обратной по отношению к функции  $Y = 4 * X$ . Придав переменной  $X$  значение, допустим, 2, получим  $Y = 8$ . Задаваясь для обратной функции значением  $Y = 8$ , получим  $X = 2$ . Для функции регрессии это несвойственно. Поэтому, если изучается стохастическая зависимость  $Y$  от  $X$ , то строят регрессию  $Y$  на  $X$ ; если же исследуется стохастическая зависимость  $X$  от  $Y$ , то строят регрессию  $X$  на  $Y$ . Например, при исследовании зависимости спроса на товар ( $Y$ ) от его стоимости ( $X$ ) строят регрессию  $Y$  на  $X$ . Если же, наоборот, изучается механизм влияния спроса на стоимость товара, то – регрессия  $X$  на  $Y$ . То есть, *хотя в данном случае зависимость обратима, функция регрессии, подлежащая определению, не обладает свойством обратимости.*

Классификация регрессионных функций осуществляется по различным признакам (рис. 1.3). На практике все приведенные на рис. 1.3 виды регрессии обычно не рассматриваются изолированно.



Рисунок 1.3 - Классификация видов регрессии

Так, применяется простая линейная и простая нелинейная регрессия, множественная линейная и множественная нелинейная, причем отдельные составляющие этих регрессий демонстрируют как положительный, так и отрицательный характер.

### 1.3 Понятие корреляции

*Корреляция в широком смысле* означает связь, существующую между различными явлениями и процессами. Связи между явлениями и процессами могут быть различными по силе. При измерении силы (тесноты) стохастической связи используется *корреляция в узком смысле*

Однако корреляция сама по себе не вскрывает причинного характера связи. С помощью корреляционного анализа нельзя узнать, какое явление принимать в качестве причины, а какое – в качестве следствия. *Корреляция лишь дает оценку силы или тесноты связи.*

Во многих случаях несложно выяснить, какие переменные представляют собой причину и что является следствием. Причем в разных исследованиях причины и следствия могут меняться местами. Так, существует корреляция между ростом производительности труда и повышением заработной платы. В общем случае рост производительности труда можно считать причиной повышения заработной платы; но, с другой стороны, повышение заработной платы может являться материальным стимулом роста производительности труда. При определении же корреляции между количеством осадков и урожайностью сельскохозяйственных культур не возникает сомнения, какие переменные принять в качестве причины и что считать следствием.

Но иногда трудно выяснить взаимоотношения между переменными, например «... между заболеванием раком легких и курением. Хотя курение в вероятностном смысле оказывает влияние на образование рака легких, нельзя, однако, утверждать, что курение является причиной заболевания» [16].

Таким образом, *установление корреляции не означает наличия причинной связи.*

Аналогично классификации регрессии, приведенной на рис. 1.3, можно классифицировать и различные виды корреляции (рис. 1.4).



Рисунок 1.4 - Классификация видов корреляции

В книге [16] приводятся следующие примеры, позволяющие лучше разобраться в сущности корреляционных зависимостей.

*Частная корреляция.* Важность использования частной корреляции определяется тем, что одновременно несколько взаимодействующих причин оказывают совместное влияние на исследуемый признак. Так, при исследовании зависимости расхода пара от объема производства на одном предприятии, изготавливающем сборные железобетонные конструкции под открытым небом, установили сильную отрицательную корреляцию, то есть, с увеличением объема производства расход пара снижался. Анализ показал, что причиной такого парадоксального вывода явилось влияние другого фактора на потребление пара, а именно, температуры воздуха. Следовательно, прежде чем определять корреляцию между расходом пара и объемом производства, надо исключить влияние температуры воздуха на потребление пара. При определении корреляции между температурой воздуха и потреблением пара также следует устранить влияние объема производства на расход пара. В результате будем иметь две частные корреляции, каждая из которых указывает «чистую» стохастическую связь между двумя переменными при элиминированном (исключенном) влиянии третьей.

*Косвенная корреляция* иллюстрируется примером из статистики дореволюционной России. Была установлена тесная корреляция между числом пожаров в стране и размером урожая. В неурожайные годы число пожаров было довольно высоким. Очевидно, плохие урожаи нельзя считать причиной пожаров в зданиях. В действительности между приведенными переменными существует только косвенная связь. Как размеры урожая, так и количество пожаров существенно зависят от третьего фактора – метеоусловий. Сильная засуха, естественно, приводит к плохому урожаю. Она же благоприятствует возникновению пожаров. Только поэтому проявляется связь между рассматриваемыми явлениями.

*Ложная корреляция.* Под ложной корреляцией понимается чисто формальная связь между явлениями, не находящая логического объяснения и

основанная лишь на количественном соотношении между ними. Часто ложная корреляция возникает при изучении экономических временных рядов, имеющих сходные тенденции. При сопоставлении таких рядов необходимо сначала исключить тенденции и только потом устанавливать корреляцию.

В статистической литературе приводится пример ложной корреляции между числом аистов, свивших гнезда в южных районах Швеции, и рождаемостью в эти же годы в Швеции. Вычисления, выполненные ради шутки, показали положительную корреляцию между этими явлениями. Этот пример подтверждает, что причинная зависимость не может быть выведена ни из какого наблюдаемого совместного изменения явлений.

Из всего вышесказанного можно сделать следующий вывод.

*Понятия регрессии и корреляции непосредственно связаны между собой: тогда, как в корреляционном анализе оценивается сила стохастической связи, в регрессионном анализе исследуется ее форма.*

Оба вида анализа служат для установления причинных соотношений между явлениями и для определения наличия или отсутствия связи.

#### **1.4 Понятие эконометрической модели и основные задачи эконометрического моделирования**

Центральной проблемой эконометрики является построение эконометрической модели и определение возможностей ее использования для описания, анализа и прогнозирования реальных экономических процессов. В настоящее время эконометрика располагает огромным разнообразием типов моделей – от больших макроэкономических моделей, включающих несколько сот, а иногда и тысяч уравнений, до малых коинтеграционных моделей, предназначенных для решения специфических проблем.

**Эконометрическая модель** состоит из регрессионных уравнений (и тождеств), которые включают *переменные и параметры*.

**Переменные** – это экономические величины, которые могут принимать любое значение из некоторого множества допустимых величин.

**Параметры** – это постоянные коэффициенты, которые связывают переменные в уравнениях. Когда параметры не известны, их можно оценить с помощью статистических методов на основе данных выборочных обследований.

Уравнения, а значит, и параметры, формируют **структуру модели**: *они указывают на характер предполагаемых соотношений между переменными.*

Можно выделить **три основных класса моделей**, которые применяются для анализа и прогноза (рис. 1.5).

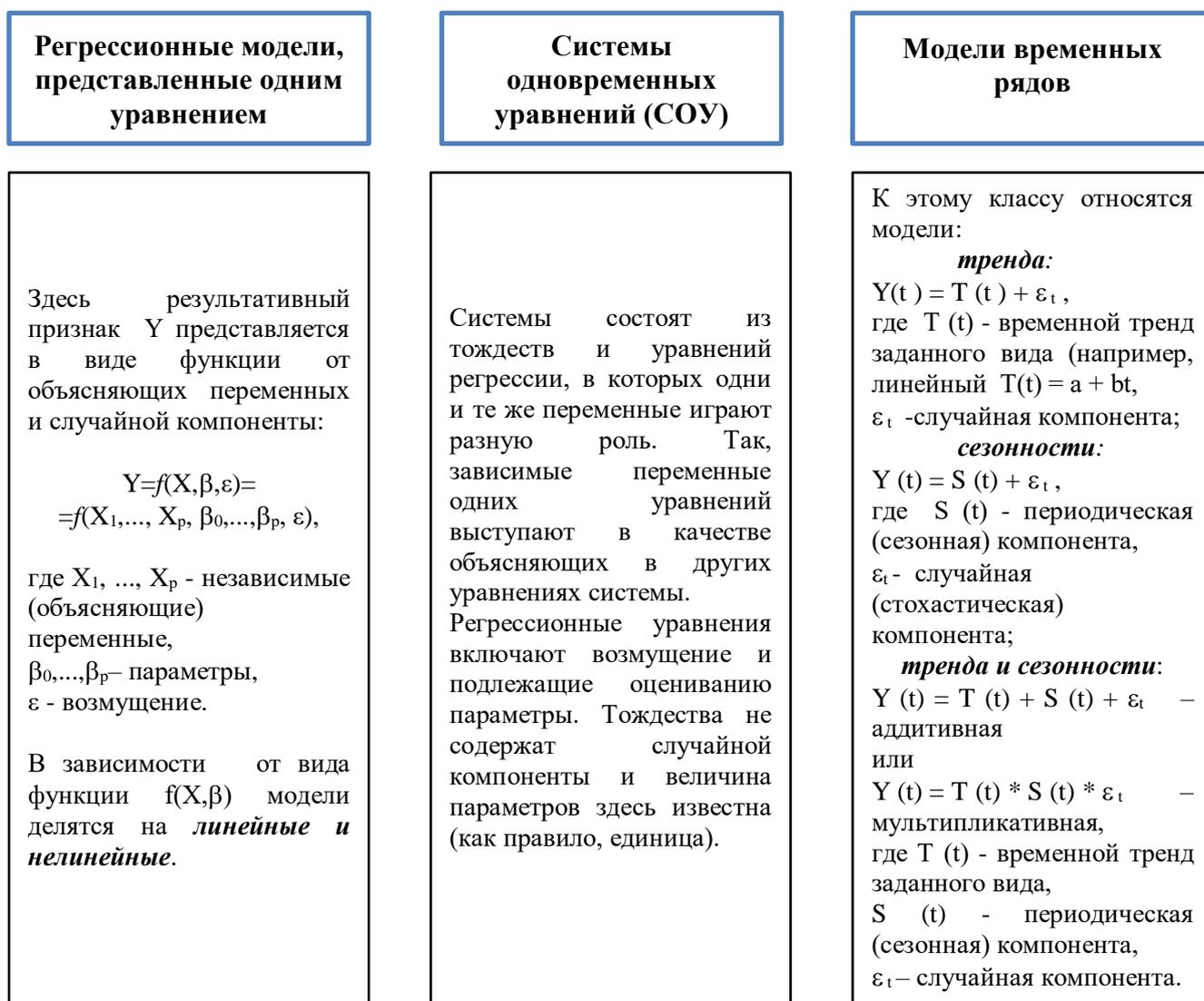


Рисунок 1.5 - Классы эконометрических моделей

Представленные на рис. 1.5 модели находят широкое применение для описания экономических процессов.

Так, модели временных рядов, к которым относится множество более сложных моделей (модели адаптивного прогноза, модели авторегрессии и скользящего среднего и др.) могут применяться, например, для изучения и прогнозирования объема продаж авиабилетов, спроса на мороженое, краткосрочного прогноза процентных ставок и т.п.

С помощью регрессионных моделей, представленных одним уравнением, можно исследовать спрос, например, на мороженое в зависимости от многих факторов: от времени, температуры воздуха, среднего уровня доходов или как зависимость зарплаты от возраста, пола, уровня образования, стажа работы и т.д. Область применения таких моделей, даже линейных, значительно шире, чем моделей временных рядов.

С помощью систем одновременных уравнений описываются сложные экономические явления и процессы, которые выступают по взаимосвязи, в частности, макроэкономические.

В основе построения и оценивания моделей лежит **формирование совокупности наблюдений**. Различают три способа формирования совокупности: временная выборка, пространственная и пространственно-временная (панельные данные) (рис. 1.6).

Совокупность наблюдений, представляющая набор исходных данных для построения эконометрической модели, должна быть качественно и количественно однородной, т. е. относиться к однотипным экономическим объектам и включать сопоставимые величины. Выводы, которые можно сделать в результате эконометрического моделирования, во многом определены качеством исходных данных – их полнотой и достоверностью.



Рисунок 1.6 - Типы данных, лежащих в основе эконометрических построений

Прежде, чем применять математико-статистический аппарат, явление следует проанализировать с содержательной точки зрения и решить, какую переменную рассматривать как *зависимую (следствие)*, или переменную, подлежащую объяснению с помощью функции регрессии, и какие переменные в ходе анализа считать *объясняющими (причины)*, *независимыми*, или *предсказывающими*.

Далее будем обозначать зависимую переменную через  $Y$ , а объясняющие переменные через  $X_j$  ( $j = 1, 2, \dots, p$ ). Переменная  $Y$ , таким образом, является функцией от переменных  $X_j$ .

С помощью функции регрессии

$$Y = f(X_1, X_2, \dots, X_p) + \varepsilon, \quad (1.1)$$

где  $f(X_1, X_2, \dots, X_p) = \hat{Y}$ ,

количественно оценивается усредненная зависимость между исследуемыми переменными.

Присутствие в модели (1.1) случайной переменной  $\varepsilon$  (или возмущения, ошибки) связано с тем, что, во-первых, модель является упрощением действительности и на самом деле есть еще и другие факторы, от которых зависит переменная  $Y$ , и, во-вторых, как правило, имеются ошибки наблюдения (любые формы наблюдения, включая документальные, содержат ошибки). **Ошибки** – одно из основных понятий эконометрии. Определение и сведение ее к минимуму является условием построения модели для прогнозирования. Ошибки делятся на систематические и случайные. Систематические ошибки имеют постоянную величину или изменяются, подчиняясь некоторой функциональной зависимости. Случайные ошибки обусловлены влиянием случайных обстоятельств при формировании показателей.

Различают линейные и нелинейные регрессии.

Линейная регрессия описывается уравнением:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon, \quad (1.2)$$

где  $\beta_0, \beta_1, \dots, \beta_p$  – неизвестные параметры модели.

Поскольку эконометрические построения основаны на методике

*Истинные значения параметров  $\beta_0, \beta_1, \dots, \beta_p$  вычислить невозможно, поскольку обычно в распоряжении исследователя находится ограниченное число наблюдений, поэтому неизвестные параметры регрессии подлежат оцениванию по определенной процедуре.*

статистического исследования, имеющей дело с выборочными данными и использующей определенные аналитические процедуры, то такое исследование обычно исходит из определенных допущений относительно свойств случайной переменной модели. При этом центральным является вопрос о методе или формуле, которая служит для **оценивания параметров модели**.

Эти формулы называются формулами оценивания в отличие от *оценок* или

численных результатов, получаемых путем расчетов по таким формулам с использованием имеющихся статистических данных.

Оценивание параметров уравнения входит в число важнейших задач регрессионного анализа. Наряду с нахождением значений оценок параметров оценивается их точность, проверяется статистическая значимость (надежность) уравнения и его коэффициентов.

Оценки параметров будем обозначать через  $b_0, b_1, \dots, b_p$  соответственно. Тогда уравнение линейной регрессии

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p \quad (1.3)$$

будет являться оценкой модели (1.2).

Многие экономические зависимости не являются линейными по своей сути и поэтому для их описания гораздо лучше подходят нелинейные соотношения.

Различают два класса нелинейных регрессий:

1) *регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам* (включает уравнения, в которых зависимая переменная линейно связана с параметрами), например,

- полиномы разных степеней 
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon,$$
  
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon;$$
- равносторонняя гипербола 
$$Y = \beta_0 + \frac{\beta_1}{X} + \varepsilon;$$

2) *регрессии, нелинейные по оцениваемым параметрам*, например,

- степенная 
$$Y = \beta_0 X^{\beta_1} \varepsilon,$$
- показательная 
$$Y = \beta_0 \beta_1^X \varepsilon,$$
- экспоненциальная 
$$Y = e^{\beta_0 + \beta_1 X} \varepsilon.$$

Путем несложных преобразований эти модели можно свести к линейному виду.

### ***Вопросы для самоконтроля***

1. Дайте определение понятия «эконометрика».
2. Что является объектом и предметом в эконометрике?
3. Какие задачи решаются в процессе эконометрического исследования?
4. Дайте определение понятия «регрессия».
5. Приведите классификацию регрессии по различным признакам.
6. Дайте определение понятия «корреляция».
7. Приведите классификацию корреляции по различным признакам.
8. Дайте определение понятия «эконометрическая модель».
9. Чем отличаются переменные от параметров?
10. Какие классы моделей могут применяться для анализа и прогнозирования?
11. Приведите классификацию переменных в эконометрических моделях.
12. В чем сущность регрессионной модели?
13. Чем модели временных рядов отличаются от системы одновременных уравнений?
14. Каковы основные проблемы эконометрического моделирования.
15. Чем модели линейного типа отличаются от нелинейных моделей?
16. Приведите примеры моделей нелинейной регрессии.

## ТЕМА 2. МОДЕЛИ ПАРНОЙ РЕГРЕССИИ

### 2.1 Построение уравнения парной регрессии методом наименьших квадратов

*Парной (простой) регрессией* называется уравнение связи двух переменных  $Y$  и  $X$  вида

$$Y = f(X, \varepsilon), \quad (2.1)$$

где  $Y$  – зависимая переменная (результативный признак);

$X$  – независимая, объясняющая переменная (признак-фактор);

$\varepsilon$  – случайная переменная (возмущение) с некоторой функцией распределения.

Линейное парное уравнений регрессии имеет вид:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (2.2)$$

оценочное уравнение:

$$Y = b_0 + b_1 X + e. \quad (2.3)$$

Уравнение (2.3) представим в виде:

$$Y = \hat{Y} + e,$$

где  $\hat{Y} = b_0 + b_1 X,$

тогда  $Y - \hat{Y} = e. \quad (2.4)$

В уравнении (2.3)  $b_0$  – *постоянная регрессии* – свободный член, определяющий точку пересечения прямой регрессии с осью ординат (рис. 2.1), т.е. показывает значение результативного признака  $Y$  при нулевой величине фактора  $X$ . Коэффициент  $b_0$  не имеет экономического смысла.

Коэффициент  $b_1$  – *коэффициент регрессии* – характеризует наклон прямой к оси абсцисс (рис. 2.1) и *показывает, на какую величину в среднем изменится  $Y$  при изменении  $X$  на одну единицу*. Знак при этом коэффициенте указывает направление связи: при  $b_1 > 0$  связь прямая (то есть, с ростом  $X$  растет и  $Y$ ); при  $b_1 < 0$  связь обратная (увеличение  $X$  способствует уменьшению  $Y$ ).

Остаточная компонента  $e$  отражает влияние на изменчивость  $Y$  других факторов, в явном виде не присутствующих в модели.

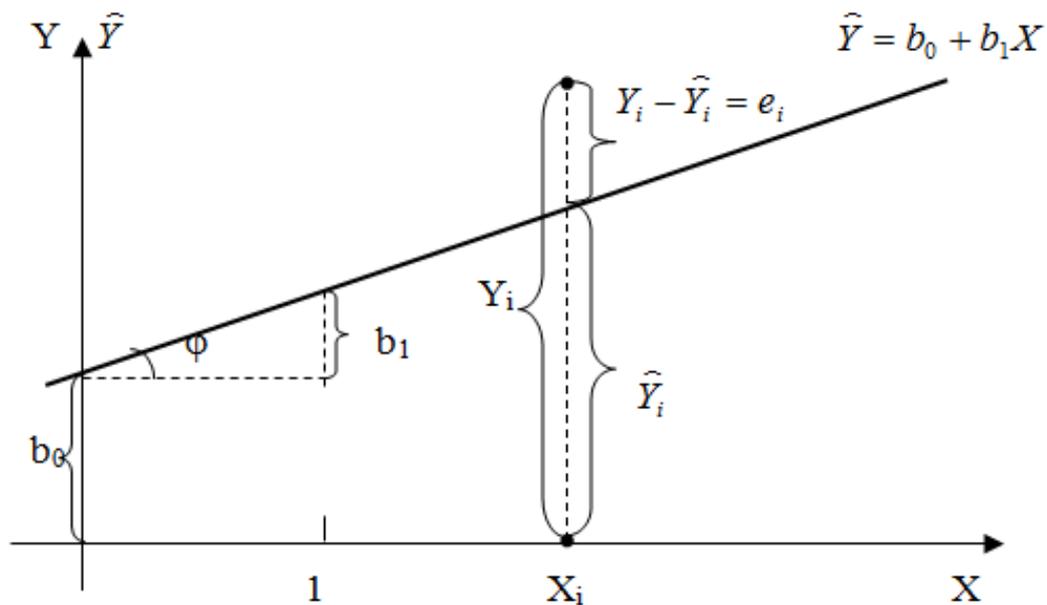


Рисунок 2.1 - Парная линейная регрессия

Один из методов оценивания параметров модели основан на принципе «*наименьших квадратов*». Этот метод представляет особый интерес, поскольку в тех случаях, когда выполняются предположения «классической» регрессионной модели, оценки, полученные с помощью метода наименьших квадратов, обладают некоторыми хорошо известными свойствами.

Сущность метода наименьших квадратов (МНК) для парной модели линейной регрессии показана на рис. 2.2.

Метод наименьших квадратов используется для оценивания линейных моделей регрессии. Но, как уже отмечалось, часто взаимосвязи экономических показателей более адекватно описываются с помощью нелинейных функций.

**Метод наименьших квадратов для парной  
линейной регрессионной модели**

Позволяет определить оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака ( $Y_i$ ) от расчетных ( $\hat{Y}_i$ ) минимальна

$$Q = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - b_0 - b_1 X_i)^2 \rightarrow \min$$

Приравняв нулю частные производные функции Q по  $b_0$  и  $b_1$  получим систему нормальных уравнений:

$$\frac{\partial Q}{\partial b_0} = -2 \sum_{i=1}^n (Y_i - b_0 - b_1 X_i) = 0$$

$$\frac{\partial Q}{\partial b_1} = -2 \sum_{i=1}^n (Y_i - b_0 - b_1 X_i) X_i = 0$$

Решение системы нормальных уравнений относительно  $b_0$  и  $b_1$  дает искомые значения оценок параметров модели

$$b_1 = \frac{n \sum_{i=1}^n Y_i X_i - \sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - \left( \sum_{i=1}^n X_i \right)^2}, \quad b_0 = \frac{\sum_{i=1}^n Y_i}{n} - b_1 \frac{\sum_{i=1}^n X_i}{n} = \bar{Y} - b_1 \bar{X}$$

Рисунок 2.2 - Метод наименьших квадратов для парной модели линейной регрессии

Нелинейные уравнения регрессии можно привести к линейному виду с помощью некоторого преобразования  $(x, y) \rightarrow (x', y')$  и применять МНК к преобразованным данным. Для наиболее применяемых функций эти вид таких преобразований представлен в табл. 2.1.

Таблица 2.1 – Приведение нелинейных уравнений регрессии к линейному виду

Вид функции регрессии	Аналитическое выражение	Линеаризующее преобразование и полученное выражение	Формулы для расчета оценок	
			$b_1$	$b_0$
Гиперболическая	$Y = \beta_0 + \beta_1 \frac{1}{X} + \varepsilon$ или $\hat{Y} = b_0 + \frac{b_1}{X}$	$X' = \frac{1}{X}, Y' = Y$ $\hat{Y} = b_0 + b_1 X'$	$b_1 = \frac{n \sum_{i=1}^n \frac{1}{X_i} Y_i - \sum_{i=1}^n \frac{1}{X_i} \sum_{i=1}^n Y_i}{n \sum_{i=1}^n \left(\frac{1}{X_i}\right)^2 - \left(\sum_{i=1}^n \frac{1}{X_i}\right)^2}$	$b_0 = \bar{Y} - b_1 \left(\overline{\frac{1}{X_i}}\right)$
Экспоненциальная	$Y = e^{\beta_0 + \beta_1 X} \varepsilon$	$\ln Y = (\beta_0 + \beta_1 X) \ln e + \ln \varepsilon$ или $\ln Y = \beta_0 + \beta_1 X + \ln \varepsilon$	$b_1 = \frac{n \sum X_i \ln Y_i - \sum X_i \sum \ln Y_i}{n \sum X_i^2 - (\sum X_i)^2}$	$b_0 = \frac{1}{n} \sum \ln Y_i - b_1 \bar{X}$
Показательная	$Y = \beta_0 \beta_1^X \varepsilon$ , ( $\beta_0 > 0, \beta_1 > 0$ )	$X' = X; Y' = \ln Y$ $\ln Y = \ln \beta_0 + X \ln \beta_1 + \ln \varepsilon$	$\ln b_1 = \frac{n \sum X_i \cdot \ln Y_i - \sum X_i \sum \ln Y_i}{n \sum X_i^2 - (\sum X_i)^2}$ $b_1 = e^{\ln b_1}$	$\ln b_0 = \frac{1}{n} \sum \ln Y_i - b_1 \frac{1}{n} \sum X_i$ $b_0 = e^{\ln b_0}$
Степенная	$Y = \beta_0 X^{\beta_1} \varepsilon$ , ( $\beta_0 > 0$ )	$X' = \ln X; Y' = \ln Y$ $\ln Y = \ln \beta_0 + \beta_1 \ln X + \ln \varepsilon$	$b_1 = \frac{n \sum \ln X_i \cdot \ln Y_i - \sum \ln X_i \sum \ln Y_i}{n \sum \ln X_i^2 - (\sum \ln X_i)^2}$	$\ln b_0 = \frac{1}{n} \sum \ln Y_i - b_1 \frac{1}{n} \sum \ln X_i$ $b_0 = e^{\ln b_0}$
Линейно-логарифмическая	$Y = \beta_0 + \beta_1 \ln X + \varepsilon$	$X' = \ln X; Y' = Y$ $Y = \beta_0 + \beta_1 X' + \varepsilon$	$b_1 = \frac{n \sum \ln X_i \cdot Y_i - \sum \ln X_i \sum Y_i}{n \sum \ln X_i^2 - (\sum \ln X_i)^2}$	$\ln b_0 = \frac{1}{n} \sum Y_i - b_1 \frac{1}{n} \sum \ln X_i$
Полиномы	$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \dots + \beta_m X^m + \varepsilon$	$X = X_1, X^2 = X_2, \dots, X^m = X_m$ $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m + \varepsilon$		

Модели *полиномов* линейны относительно параметров и могут быть сведены к линейной регрессионной модели простой заменой переменных.

Как показывает опыт исследователей, среди нелинейной полиномиальной регрессии чаще всего используется квадратичная парабола, иногда – полином третьей степени. Парабола второй степени целесообразна к применению, если для определенного интервала значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная – на прямую. Если же исходные данные не обнаруживают изменения направленности связи, то параметры квадратичной параболы становятся трудно интерпретируемыми, а форма связи часто заменяется другими нелинейными моделями, например, степенной. Полиномы высоких степеней практически не используются.

*Гиперболическая регрессия* применяется в тех случаях, когда неограниченное увеличение  $X$  асимптотически приближает зависимую переменную  $Y$  к некоторому пределу (здесь к  $\beta_0$ ).

*Экспоненциальная регрессия (показательная функция)* относится к моделям, нелинейным относительно оцениваемых параметров, так как включают параметры  $\beta_0$  и  $\beta_1$  не аддитивно. Но путем логарифмирования они приводятся к линейному виду (см. табл. 2.1).

Практическое применение экспоненты (показательной кривой) возможно, если все значения зависимой переменной  $Y$  положительны. Поэтому, если исследуется, например, финансовый результат деятельности предприятий, среди которых наряду с прибыльными есть и убыточные, то данная функция не может быть использована.

*Степенная функция* тоже сводится к линейному виду путем логарифмирования. Ее часто используют для описания экономических процессов из-за того, что коэффициент регрессии здесь является коэффициентом эластичности, т.е. он показывает, что с изменением  $X$  на 1% величина  $Y$  изменяется в среднем на  $\beta_1$  %.

*Линейно-логарифмическая модель* обычно используется, когда необходимо исследовать влияние процентного изменения объясняющей переменной  $X$  на абсолютное изменение зависимой переменной. Здесь коэффициент  $\beta_1$  определяет изменение переменной  $Y$  при единичном приросте  $X$  (например, на 1%), т.е. характеризует отношение абсолютного изменения  $Y$  к относительному изменению  $X$ . Например, если  $Y = \text{ВНП}$  (валовой национальный продукт), а  $X = M$  (денежная масса), то получаем следующую формулу:  $\text{ВНП} = \beta_0 + \beta_1 \ln M + \varepsilon$ . Из этой формулы следует, что если увеличить предложение денег  $M$  на 1%, то ВНП в среднем вырастет на  $0,01\beta_1$ .

Таким образом, в моделях, нелинейных по оцениваемым параметрам, но приводимых к линейному виду, *МНК применяется не к исходным данным результативного и факторного признаков, а к их преобразованным величинам.*

## 2.2 Оценка тесноты связи и полноты модели

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции  $r_{xy}$  для линейной регрессии ( $-1 \leq r_{xy} \leq 1$ )

$$r_{xy} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{(n \sum X_i^2 - (\sum X_i)^2)(n \sum Y_i^2 - (\sum Y_i)^2)}} = \frac{\overline{XY} - \bar{Y} \cdot \bar{X}}{\sigma_X \sigma_Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = b_1 \frac{\sigma_X}{\sigma_Y} \quad (2.5)$$

и индекс корреляции  $\rho_{xy}$  для нелинейной регрессии ( $0 \leq \rho_{xy} \leq 1$ )

$$\rho_{xy} = \sqrt{1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}} \quad (2.6)$$

Для проверки их статистической значимости рассчитывается критерий

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}, \quad (2.7)$$

который сравнивается с табличным значением при выбранной вероятности и  $(n - 2)$  степенях свободы. Превышение вычисленной  $t$ -статистики над табличным значением служит свидетельством статистической надежности полученного коэффициента.

Долю общей вариации анализируемой зависимой переменной  $Y$ , объясняемую регрессией (изменением факторных переменных), в общей дисперсии результативного признака  $Y$  характеризует **коэффициент детерминации** (для линейной регрессии) или **индекс детерминации** (для нелинейной регрессии):

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\sum(Y_i - \hat{Y})^2}{\sum(Y_i - \bar{Y})^2} . \quad (2.8)$$

**Коэффициент детерминации меняется от нуля (когда факторы  $X$  не влияют на изменчивость  $Y$ ) до единицы (когда изменение  $Y$  полностью объясняется изменением  $X$ ). Таким образом, коэффициент детерминации характеризует «полноту» модели.**

Для случая парной регрессии  $R^2$  равняется квадрату коэффициента корреляции переменных  $Y$  и  $X$  ( $r_{YX}^2$ ).

При сравнении моделей с различным числом объясняющих переменных или моделей с различными зависимыми переменными используется **скорректированный коэффициент детерминации**, который рассчитывается по формуле (2.9):

$$\bar{R}^2 = 1 - \left[ (1 - R^2) \frac{(n-1)}{(n-p-1)} \right] \quad (2.9)$$

Величина скорректированного коэффициента детерминации увеличится только в случае, когда вновь вводимая объясняющая переменная вносит весомый вклад в описание изменчивости исследуемого показателя.

## 2.3 Проверка статистической значимости (надежности) результатов оценивания

Построенная МНК модель основывается на выборочных статистических данных, то есть, результаты оценивания будут отличаться от соответствующих величин, рассчитанных по генеральной совокупности. Поэтому важно проверить, насколько представительны данные оценки относительно истинных параметров модели. Это достигается проверкой статистической надежности (значимости) коэффициентов регрессии и близости расположения фактических данных к рассчитанной линии регрессии. Если не проводить такой проверки, то можно прийти к ложным заключениям о существовании связи там, где ее нет

*Статистическая значимость результата* представляет собой оцененную меру уверенности в его "истинности" (в смысле "репрезентативности выборки").

Для характеристики статистической значимости вводится понятие уровня статистической значимости  $\alpha$ .

$\alpha$ -уровень представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю генеральную совокупность. Например,  $\alpha = 0,05$  показывает, что имеется 5% вероятность, что найденная в выборке связь между переменными является лишь случайной особенностью данной выборки.

Величина  $(1 - \alpha)$  характеризует вероятность надежности полученного результата в смысле его распространения на всю генеральную совокупность.

### 2.3.1 Проверка статистической значимости коэффициентов регрессии

Для проверки статистической значимости рассчитываются *t-критерии*:

$$t_j = \frac{|b_j|}{s_{b_j}}, \quad j = \overline{0, p}, \quad (2.10)$$

где  $s_{b_j}$  – стандартные ошибки оценок:

$$s_{b_0} = s_e \sqrt{\frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2}}, \quad s_{b_1} = s_e \sqrt{\frac{1}{\sum (X_i - \bar{X})^2}}, \quad (2.11)$$

а  $Se$  представляет меру разброса опытных данных  $(X_i, Y_i)$  вокруг значений регрессии – *стандартную ошибку регрессии (остатков)*:

$$s_e = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2}} = s_{\hat{Y}}. \quad (2.12)$$

Вычисленное значение t-критерия сравнивают с критическим, взятым из таблиц распределения Стьюдента при уровне значимости  $\alpha\%$  и числе степеней свободы  $\nu = n - 2$ . Обычно статистическую значимость проверяют при уровне доверительной вероятности 95% или 99%. Это означает, что существует вероятность 95% или 99%, что значения  $b_0$  и  $b_1$  не случайны и являются значимыми оценками соответствующих параметров.

Коэффициенты регрессии значимы, если вычисленное значение t-критерия больше величины, указанной в таблице t-распределения Стьюдента.

### 2.3.2 Проверка статистической значимости уравнения регрессии

Оценка адекватности выбранной модели эмпирическим данным осуществляется на основе коэффициента детерминации  $R^2$  с помощью F-теста. Выполняется сравнение фактического  $F_{\text{факт}}$  и критического (табличного)  $F_{\text{табл}}$  значений F-критерия Фишера. Расчет F-критерия для парной модели выполняется следующим образом:

$$F_{\text{факт}} = \frac{\frac{R^2}{1-R^2}}{\frac{1}{n-2}}. \quad (2.13)$$

Табличное значение F-критерия для вероятности 0,95 (или 0,99) и числа степеней свободы  $\nu_1 = 1, \nu_2 = n - 2$  сравнивается с вычисленным.

При выполнении неравенства  $F > F_{\text{табл}}$  отвергается нулевая гипотеза о том, что истинное значение коэффициента детерминации равно нулю; это дает основание считать, что модель адекватна исследуемому процессу.

### 2.3.3 Интервальное оценивание параметров линейной регрессии

Расчитанные «точечные» значения оценок (коэффициенты  $b_0, b_1$ ) являются приближенными, полученными на основе имеющихся выборочных данных. Для определения того, насколько истинные значения параметров могут отличаться от вычисленных оценок, осуществляется построение *доверительных интервалов*, задающих пределы, в которых должны находиться точные значения параметров с заданной доверительной вероятностью (степенью уверенности, соответствующей заданному уровню значимости  $\alpha$ ).

Для построения доверительных интервалов определяем предельную ошибку  $\Delta$  для каждого параметра  $\beta_0$  и  $\beta_1$  уравнения линейной регрессии:

$$\Delta_0 = t_{\alpha, v} S_{b_0}, \quad \Delta_1 = t_{\alpha, v} S_{b_1} \quad (2.14)$$

Величина  $t_{\alpha, v}$  представляет собой табличное значение  $t$ -критерия Стьюдента при степени свободы  $v = n - 2$  и заданном уровне значимости  $\alpha$  (двустороннее оценивание).

Формулы для расчета доверительных интервалов имеют следующий вид:

$$b_j - \Delta_j \leq \beta_j \leq b_j + \Delta_j,$$

или 
$$b_j - t_{\alpha, v} s_j < \beta_j < b_j + t_{\alpha, v} s_j. \quad (2.15)$$

Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, то есть, его оценка считается статистически не значимой.

## 2.4 Точечный и интервальный прогноз по уравнению линейной регрессии

**Точечный прогноз – это величина, которая получается путем подстановки известного значения ( $X_0$ ) независимой переменной в оцененное уравнение регрессии  $\hat{Y}_0 = b_0 + b_1 X_0$**

Однако точечный прогноз считается нереалистичным, поэтому он дополняется интервальным прогнозом.

**Интервальный прогноз – это доверительный интервал, в котором с определенной вероятностью может находиться фактическое значение прогнозируемой величины**

Существуют два вида интервальных прогнозов: для математических ожиданий (для функции регрессии) и для индивидуальных значений зависимой переменной.

### *Доверительный интервал для функции регрессии*

С вероятностью  $(1 - \alpha)$  можно утверждать, что уравнение регрессии для средних значений (математических ожиданий) зависимой переменной  $M_X(Y)$  будет находиться в интервале:

$$\hat{Y}_0 - t_{\alpha;v} \cdot s_p \leq M_X(Y) \leq \hat{Y}_0 + t_{\alpha;v} \cdot s_p \quad (2.16)$$

где  $s_p = s_{\hat{Y}} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$  – стандартная ошибка предсказываемого среднего

значения  $Y$  при заданном значении  $X_0$ , которая характеризует ошибку положения линии регрессии.

$s_{\hat{Y}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}}$  – стандартная ошибка остатков (модели).

### ***Доверительный интервал для индивидуальных значений зависимой переменной***

Построенная доверительная область для  $M_x(Y)$  определяет местоположение модельной линии регрессии (т.е. условного математического ожидания), но не отдельных возможных значений зависимой переменной, которые отклоняются от средней. Поэтому ошибка предсказываемого индивидуального значения  $Y$  должна включать не только стандартную ошибку  $s_p$ , но и случайную ошибку  $s_Y$ .

Средняя ошибка прогнозируемого индивидуального значения  $Y$  при  $X = X_0$  равна:

$$s_p = s_{\hat{Y}} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}. \quad (2.17)$$

а соответствующий доверительный интервал для прогнозов индивидуальных значений  $Y_0^*$  будет определяться по формуле:

$$\hat{Y}_0 - t_{1-\alpha;v} \cdot s_{\hat{Y}} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}} \leq Y_0^* \leq \hat{Y}_0 + t_{1-\alpha;v} \cdot s_{\hat{Y}} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}. \quad (2.18)$$

## **2.5 Коэффициент эластичности**

Коэффициенты регрессии (в рассматриваемом случае это коэффициент  $b_1$ ) нельзя использовать для непосредственной оценки влияния факторов на результативный признак из-за различия единиц измерения исследуемых показателей. Для этих целей вычисляются *коэффициенты эластичности*, которые в экономических исследованиях находят широкое применение.

Коэффициент эластичности для модели парной регрессии рассчитывается по формуле (2.19):

$$E_{Y/x} = \frac{dY}{dX} \cdot \frac{X}{Y} \quad (2.19)$$

и показывает, насколько процентов изменится результат  $Y$  при изменении фактора  $X$  на 1% от своего номинального значения.

Для линейной регрессии коэффициент эластичности равен

$$E_{Y/x} = b_1 \cdot \frac{X}{Y} \quad (2.20)$$

и зависит от  $X$  и  $Y$ , поэтому рассчитывают средний коэффициент эластичности:

$$\bar{E}_{Y/x} = b_1 \cdot \frac{\bar{X}}{\bar{Y}}. \quad (2.21)$$

Средний коэффициент эластичности показывает, на сколько процентов в среднем по совокупности изменится результат  $Y$  от своей величины при изменении фактора  $X$  на 1 % от своего значения.

## 2.6 Выбор наилучшей функции регрессии

Если форма связи между экономическими показателями неизвестна, то следует построить несколько функций регрессии (на основе предположений относительно характера связи между показателями) и выбрать из них наилучшую.

При прочих равных условиях критериями выбора наилучшей функции являются следующие:

- 1) статистическая значимость оценок параметров модели;
- 2) статистическая значимость модели в целом;
- 3) выполнение требований Гаусса - Маркова, предъявляемых к случайным остаткам модели (основных предпосылок), в первую очередь, постоянство дисперсии и независимость друг от друга;

4) минимальная доля остаточной дисперсии в общей дисперсии, т.е. максимальная величина коэффициента детерминации  $R^2$ . Если модели регрессии содержат разное количество параметров, вместо  $R^2$  следует сравнивать скорректированные коэффициенты детерминации  $R^2_{\text{скор}}$ .

Аналогично линейному коэффициенту детерминации (обозначим его  $r^2$ ) для нелинейных функций вычисляется индекс детерминации ( $R^2$ ). Чем больше кривизна линии регрессии, тем величина коэффициента детерминации  $r^2$  меньше индекса детерминации  $R^2$ . Близость этих показателей означает, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию. Практически, если величина  $(R^2 - r^2)$  не превышает 0,1, то предположение о линейной зависимости считается оправданным. В противном случае проводится оценка существенности различия показателей детерминации, вычисленных по одним и тем же данным, с использованием  $t$ -критерия Стьюдента:

$$t = \frac{R^2 - r^2}{m_{|R-r|}}, \quad (2.22)$$

где  $m_{|R-r|}$  – ошибка разности между  $R^2$  и  $r^2$ , определяемая по формуле (2.23):

$$m_{|R-r|} = 2 \cdot \sqrt{\frac{(R^2 - r^2) - (R^2 - r^2)^2 \cdot (2 - (R^2 + r^2))}{n}}. \quad (2.23)$$

Если  $t > t_{\text{табл}} = t_{(\alpha; n-p-1)}$ , то различия между показателями корреляции существенны и замена нелинейной регрессии уравнением линейной функции невозможна. Практически, если величина  $t < 2$ , то различия между  $R$  и  $r$  несущественны, и, следовательно, возможно применение линейной регрессии, даже если есть предположения о некоторой нелинейности рассматриваемых соотношений факторов и результата.

Следует отметить, что сопоставление функций по коэффициентам детерминации можно проводить только в том случае, если зависимая переменная представлена в моделях в одной и той же форме. Например, в линейной,

полулогарифмической, полиномиальной функциях зависимая переменная представлена в *исходной форме*:

$Y = \beta_0 + \beta_1 x + \varepsilon$  – линейная функция,

$Y = \beta_0 + \beta_1 \ln x + \varepsilon$  – полулогарифмическая функция,

$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$  – полином второй степени.

А в степенной или показательной функциях линейаризация связана с преобразованием зависимой переменной:

$\ln Y = \ln \beta_0 + \beta_1 \ln x + \ln \varepsilon$  – линейаризация степенной функции,

$\ln Y = \ln \beta_0 + x \ln \beta_1 + \ln \varepsilon$  – линейаризация показательной функции.

И коэффициент детерминации рассчитывается для значений *преобразованной переменной*. Так, при логарифмировании зависимой переменной (например, для степенной или показательной функций) коэффициент детерминации определится по формуле (2.24):

$$R^2 = 1 - \frac{\sum (\ln Y - \ln \hat{Y})^2}{\sum (\ln Y - \overline{\ln Y})^2}, \quad (2.24)$$

при преобразовании зависимой переменной в обратной функции

$$R^2 = 1 - \frac{\sum (1/Y - 1/\hat{Y})^2}{\sum (1/Y - \overline{1/Y})^2}, \quad (2.25)$$

где  $Y = \frac{1}{\beta_0 + \beta_1 x_1 + \varepsilon}, \quad \frac{1}{Y} = \beta_0 + \beta_1 x_1 + \varepsilon.$

**Очевидно, что показатели детерминации, рассчитанные по исходным и преобразованным значениям зависимой переменной, сравнивать нельзя**

Для сравнения моделей регрессии можно использовать *среднюю ошибку аппроксимации* – среднее относительное отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{Y_i - \hat{Y}}{Y_i} \right| \cdot 100\%. \quad (2.26)$$

Построенное уравнение регрессии считается удовлетворительным, если значение  $\bar{A}$  не превышает 10–12 %.

## 2.7 Построение эконометрической модели с использованием EXCEL

### 2.7.1 Функции Excel для вычисления коэффициентов парной линейной регрессии

**Функция ОТРЕЗОК.** Вычисляет коэффициент  $b_0$  и обращение имеет вид ОТРЕЗОК(диапазон значений  $y$ ; диапазон значений  $x$ ).

**Функция НАКЛОН.** Вычисляет коэффициент  $b_1$  и обращение имеет вид НАКЛОН(диапазон значений  $y$ ; диапазон значений  $x$ ).

**Функция ПРЕДСКАЗ.** Вычисляет значение линейной парной регрессии при заданном значении независимой переменной (обозначена через  $z$ ) и обращение имеет вид ПРЕДСКАЗ( $z$ ; диапазон значений  $y$ ; диапазон значений  $x$ ).

**Функция СТОШУХ.** Вычисляет оценку  $S$  для среднеквадратического отклонения  $\sigma$  возмущений  $\varepsilon_i$  и обращение имеет вид (YX – латинские буквы): СТОШУХ(диапазон значений  $y$ ; диапазон значений  $x$ ).

Коэффициенты уравнения линейной регрессии по пространственной выборке можно вычислить, используя функции Excel. Фрагмент документа Excel, вычисляющего требуемые величины приведен на рис. 2.3.

	A	B	C	D	E	F
1	Исходные данные		=ПРЕДСКАЗ(A3:\$B\$3:\$B\$12;\$A\$3:\$A\$12)			
2	$x_i$	$y_i$	$\hat{y}_i$			
3	8	5	5,377			
4	11	10	8,426			
5	12	10	9,443			
6	9	7	6,393			
7	8	5	5,377			
8	8	6	5,377			
9	9	6	6,393			
10	9	5	6,393			
11	8	6	5,377			
12	12	8	9,443			
13						
14						
15		$b_0$	-2,754	=ОТРЕЗОК(B3:B12;A3:A12)		
16		$b_1$	1,016	=НАКЛОН(B3:B12;A3:A12)		
17		$S$	1,024	=СТОШУХ(B3:B12;A3:A12)		

Рисунок 2.3- Использование функций Excel

### 2.7.2 Применение стандартной программы «ЛИНЕЙН» для оценивания параметров линейной регрессии

Встроенная статистическая функция **ЛИНЕЙН** оценивает коэффициенты линейной регрессии:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p.$$

Порядок вычислений следующий:

- 1) введите исходные данные;
- 2) выделите область пустых ячеек, включающую 5 строк и  $(p + 1)$  столбцов (где  $p$  – количество независимых переменных) для вывода результатов регрессионной статистики; активизируйте **Мастер функций**: щелкните по кнопке **Вставка функции**  $f_x$
- 3) в раскрывшемся окне выберите Категорию **Статистические**, Функцию – **ЛИНЕЙН**. Щелкните по кнопке **ОК**;
- 4) заполните аргументы функции (рис. 2.4):

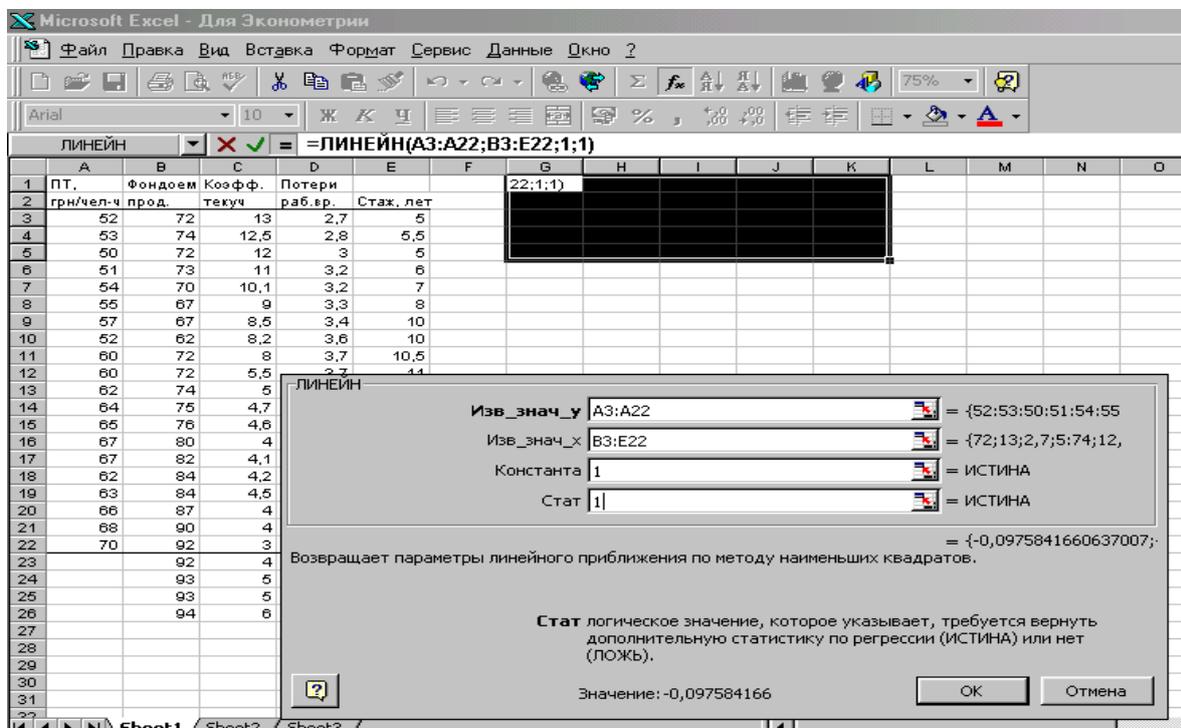


Рисунок 2.4 - Мастер функций. Работа с функцией **ЛИНЕЙН**

- *Известные значения Y* – диапазон, содержащий данные, характеризующие результативный признак;
- *Известные значения X* – диапазон, содержащий данные, описывающие все независимые переменные;
- *Константа* – логическое значение, указывающее на наличие или отсутствие свободного члена в уравнении; если *Константа* = 1, то свободный член присутствует, если *Константа* = 0, то свободный член = 0;

- *Статистика* – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. Если *Статистика* = 1, то дополнительная информация выводится, если *Статистика* = 0, то выводятся только оценки параметров уравнения.

Щелкните по кнопке **ОК**;

- 5) в левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу <F2>, а затем – на комбинацию клавиш <Ctrl>+<Shift>+<Enter>. Дополнительная регрессионная статистика будет выводиться в виде, показанном в табл. 2.2.

Таблица 2.2 – Результаты, выводимые при использовании функции

### ЛИНЕЙН

Значение $b_p$	Значение $b_{p-1}$	...	Значение $b_2$	Значение $b_1$	Значение $b_0$
Стандартная ошибка оценки $b_p$ ( $s_{b_p}$ )	Стандартная ошибка оценки $b_{p-1}$ ( $s_{b_{p-1}}$ )	...	Стандартная ошибка оценки $b_2$ ( $s_{b_2}$ )	Стандартная ошибка оценки $b_1$ ( $s_{b_1}$ )	Стандартная ошибка оценки $b_0$ ( $s_{b_0}$ )
Коэффициент детерминации $R^2$	Стандартная ошибка модели (остатков) $S_e$	#Н/Д	#Н/Д	#Н/Д	#Н/Д
F-статистика	Число степеней свободы $n-(p+1)$	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Регрессионная сумма квадратов	Остаточная сумма квадратов	#Н/Д	#Н/Д	#Н/Д	#Н/Д

Функция **ЛИНЕЙН** используется и для расчета оценок параметров нелинейных моделей, которые с помощью преобразования могут быть сведены к линейному виду. Например, степенная функция  $\hat{Y} = b_0 X_1^{b_1} X_2^{b_2} X_3^{b_3} X_4^{b_4}$  путем логарифмирования превращается в линейную по параметрам функцию:

$$\ln \hat{Y} = \ln b_0 + b_1 \ln X_1 + b_2 \ln X_2 + b_3 \ln X_3 + b_4 \ln X_4,$$

или

$$Z = A + b_1 Z_1 + b_2 Z_2 + b_3 Z_3 + b_4 Z_4,$$

где  $Z = \ln \hat{Y}$ ,  $z_j = \ln X_j$  ( $j = 1, 2, 3, 4$ );  $A = \ln b_0$ , т.е.  $b_0 = \exp(A) = e^A$ , где  $e$  – основание натурального логарифма.

Из способа преобразования видно, что для вычисления коэффициентов степенной функции с помощью **ЛИНЕЙН** следует в строки *Известные значения Y* и *Известные значения X* окна рассматриваемой функции вводить логарифмы исходных значений  $Y$  и  $X$ .

### Вопросы для самоконтроля

1. В чем состоит метод наименьших квадратов? Как он реализуется при оценивании параметров линейной модели наблюдений? Что такое система нормальных уравнений?

2. Чему равна алгебраическая сумма остатков, полученных при оценивании параметров линейной модели наблюдений?

3. В каком соотношении находятся знак оценки наименьших квадратов углового коэффициента модели прямолинейной связи  $\hat{Y} = b_0 + b_1X$  и знак выборочной ковариации  $\text{Cov}(x, y)$ ?

4. Зависит ли значение оценки наименьших квадратов углового коэффициента модели прямолинейной связи  $\hat{Y} = b_0 + b_1X$  от выбора единиц измерения переменных  $x$  и  $y$ ?

5. Каково соотношение между полной, объясненной моделью и остаточной суммами квадратов, получаемыми в результате оценивания методом наименьших квадратов линейной модели наблюдений?

6. Что такое коэффициент детерминации? Для какой цели он предназначен? В каких границах он изменяется и когда достигает своих граничных значений?

7. Какие переменные называются объясняющими, а какие — объясняемыми?

8. Что можно сказать о корреляционной связи между остатками и прогнозными значениями объясняемой переменной?

## ТЕМА 3. МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

### 3.1. Оценивание параметров многофакторной регрессии методом наименьших квадратов

Истинная взаимосвязь между результирующим показателем (зависимой переменной)  $Y$  и различными объясняющими переменными  $X_j$  выражается так:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (3.1)$$

или в матричном виде

$$Y = X\beta + \varepsilon,$$

где  $Y$  – вектор значений зависимой переменной;

$X$  – матрица значений независимой переменной размерностью  $(n \times (p+1))$ ;

$n$  – количество наблюдений;

$(p+1)$  – количество независимых переменных, включая значений фиктивной переменной  $X_0$ , тождественно равной единице (или число оцениваемых параметров);

$\beta$  – вектор параметров модели;

$\varepsilon$  – вектор возмущения:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix}; \quad \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ 1 & X_{31} & X_{32} & \dots & X_{3p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix}; \quad \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}; \quad \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \quad (3.2)$$

Однако, как и в случае простой линейной регрессии, мы не знаем истинную зависимость и вынуждены находить оценки:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p + \varepsilon. \quad (3.3)$$

Многофакторная оценочная регрессионная модель в матричном виде представляется следующим образом:

$$\hat{Y} = XB, \quad (3.4)$$

Применение метода наименьших квадратов для оценивания параметров множественной линейной регрессии показано на рис. 3.1.



Рисунок 3.1 - Метод наименьших квадратов для множественной модели линейной регрессии

Для оценивания параметров множественной линейной регрессии методом наименьших квадратов применяют следующие формулы:

$$B = (X'X)^{-1} X'Y, \quad (3.5)$$

где

$$X'X = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & X_{31} & \dots & X_{n1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{n2} \\ X_{13} & X_{23} & X_{33} & \dots & X_{n3} \end{pmatrix} \cdot \begin{pmatrix} 1 & X_{11} & X_{12} & X_{13} \\ 1 & X_{21} & X_{22} & X_{23} \\ 1 & X_{31} & X_{32} & X_{33} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i2} & \sum_{i=1}^n X_{i3} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \sum_{i=1}^n X_{i1}X_{i2} & \sum_{i=1}^n X_{i1}X_{i3} \\ \sum_{i=1}^n X_{i2} & \sum_{i=1}^n X_{i2}X_{i1} & \sum_{i=1}^n X_{i2}^2 & \sum_{i=1}^n X_{i2}X_{i3} \\ \sum_{i=1}^n X_{i3} & \sum_{i=1}^n X_{i3}X_{i1} & \sum_{i=1}^n X_{i3}X_{i2} & \sum_{i=1}^n X_{i3}^2 \end{pmatrix};$$

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & X_{31} & \dots & X_{n1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{n2} \\ X_{13} & X_{23} & X_{33} & \dots & X_{n3} \end{pmatrix} \cdot \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_i X_{i1} \\ \sum_{i=1}^n Y_i X_{i2} \\ \sum_{i=1}^n Y_i X_{i3} \end{pmatrix}$$

При решении этой задачи удобно воспользоваться электронными таблицами, где имеются программы обращения и умножения матриц.

Вычисления следует производить в следующем порядке.

1. Найти произведение матриц  $X'X$ . (функция МУМНОЖ)
2. Найти матрицу, обратную к матрице  $X'X$ . (функция МОБР)
3. Найти произведение матриц  $X'Y$ . (функция МУМНОЖ)
4. Найти произведение матриц  $(X'X)^{-1}$  и  $X'Y$ , которое дает искомые оценки параметров модели.
5. При оценивании параметров нелинейной модели методом наименьших квадратов модель следует привести к линейному виду с помощью замены переменных:  $Z = X^2$ . Получаем:

$$Y = b_0 + b_1X + b_2Z. \quad (3.6)$$

Теперь можно использовать общую формулу (3.5) расчета оценок параметров множественной линейной модели с учетом наличия двух переменных  $X$  и  $Z$ , предварительно преобразовав исходную информацию: введя в рассмотрение столбец квадратов значений переменной  $X$ .

Степенную модель следует привести к линейному виду путем логарифмирования:

$$\ln Y = \ln b_0 + b_1 \ln X_1 + b_2 \ln X_2 + b_3 \ln X_3, \quad (3.7)$$

т.е. прежде, чем применять общий подход (3.5) к расчету коэффициентов уравнения, следует найти логарифмы значений, описывающих переменные задачи. Это легко сделать, воспользовавшись соответствующей программой в EXCEL. Необходимо обратить внимание на следующее обстоятельство: по формуле (3.7) получим величину коэффициентов  $b_1, b_2, b_3$  в натуральном виде, а  $b_0$  – в виде логарифма. Для получения  $b_0$  следует найти антилогарифм, обратившись к функции: =exp( $b_0$ ) в EXCEL.

Расчет оценок параметров моделей (3.6) и (3.7) после указанных преобразований рекомендуется осуществлять, прибегнув к функции ЛИНЕЙН в EXCEL.

*Вычисление стандартной ошибки остатков (уравнения)* осуществляется по формуле (3.8):

$$S_e = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - (p + 1)}}, \quad (3.8)$$

где  $p$  – количество независимых переменных (факторов  $X$ ). Расчетные значения  $\hat{Y}$  легко найти в EXCEL с помощью функции ТЕНДЕНЦИЯ, а разности между фактическими и расчетными значениями функции – копированием соответствующей формулы.

Лучшее уравнение надо выбирать по минимальной стандартной ошибке, но следует иметь в виду, что если расчет оценок параметров осуществлялся в EXCEL с помощью функции ЛИНЕЙН по преобразованным данным, то принимать во

внимание стандартные ошибки, вычисленные компьютером и выведенные в общем массиве информации, нельзя, т.к. их величина будет несопоставима. Поэтому для полученных моделей эти ошибки необходимо считать по формуле (3.8), предварительно вычислив  $\hat{Y}$ .

*Стандартные ошибки оценок параметров* модели можно найти по формуле (3.9):

$$S_{b_j} = S_e \sqrt{X^{(jj)}}, \quad (3.9)$$

где  $X^{(jj)}$  – диагональные элементы матрицы  $(X'X)^{-1}$ .

Для проверки *статистической значимости коэффициентов регрессии* вычисленные t-статистики

$$t_j = \frac{b_j}{S_{b_j}} \quad (3.10)$$

сравниваются с табличными при доверительной вероятности 0,95 и числе степеней свободы  $n-(p+1)$ . При выполнении неравенства  $|t_j| > t_{\text{табл}}$  делают вывод о статистической надежности соответствующего коэффициента регрессии.

*Множественный коэффициент детерминации* определяется так:

$$R^2 = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}. \quad (3.11)$$

Для *проверки адекватности* выбранной модели эмпирическим данным воспользуемся F-критерием Фишера:

$$F = \frac{R^2(n-p-1)}{(1-R^2)p}. \quad (3.12)$$

Табличное значение F-критерия для вероятности 0,95 и числа степеней свободы  $\nu_1 = p$ ,  $\nu_2 = n - (p+1)$  сравниваем с вычисленным; выполнение неравенства  $F > F_{\text{табл}}$  дает основание считать, что модель адекватна исследуемому процессу.

*Доверительные интервалы* для параметров линейной регрессии:

$$b_j \pm t_{n-p-1, \alpha} S_{bj}. \quad (3.13)$$

*Доверительный интервал для прогнозных значений:*

- условного математического ожидания

$$\hat{Y}_0 - t_{\alpha, \nu} s_{\hat{Y}} \sqrt{\mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0} \leq M_X(Y) \leq \hat{Y}_0 + t_{\alpha, \nu} s_{\hat{Y}} \sqrt{\mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0}, \quad (3.14)$$

- индивидуальных значений зависимой переменной:

$$\hat{Y}_0 - t_{\alpha, \nu} s_{\hat{Y}} \sqrt{1 + \mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0} \leq Y_0^* \leq \hat{Y}_0 + t_{\alpha, \nu} s_{\hat{Y}} \sqrt{1 + \mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0}, \quad (3.15)$$

где  $\mathbf{X}_0$  – матрица ожидаемых значений объясняющих переменных, т.е. предполагается, что объясняющие переменные  $X_1, X_2, \dots, X_p$  приняли значения, задаваемые вектором:  $X'_0 = (1 \ X_{10} \ X_{20} \ \dots \ X_{p0})$ ;

$\hat{Y}_0$  – точечный прогноз, который получается подстановкой в модель значений независимых переменных  $X'_0 = (1 \ X_{10} \ X_{20} \ \dots \ X_{p0})$ ;

$$s_p = s_{\hat{Y}} \sqrt{\mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0} \quad \text{и} \quad s_p = s_{\hat{Y}} \sqrt{1 + \mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0} -$$

стандартная ошибка прогноза для условного математического ожидания и индивидуальных значений зависимой переменной соответственно;

$$s_e = \sqrt{\frac{\sum (y - \hat{Y})^2}{n - (p + 1)}} - \text{стандартная ошибка остатков (уравнения);}$$

$t_{\alpha, \nu}$  – критическое (табличное) значение t-критерия при уровне значимости  $\alpha$  и  $n - (p + 1)$  степенях свободы.

Как отмечалось ранее, использовать коэффициенты регрессии для сопоставления степени влияния факторов на изменчивость исследуемого показателя нельзя из-за различия единиц измерения и разной степени колеблемости.

Ранжировать факторы  $X_j$  по степени их влияния на показатель  $Y$  можно с помощью *коэффициентов эластичности  $E_j$*

$$E_{x_j} = \frac{\partial Y}{\partial X_j} \frac{\bar{X}_j}{\bar{Y}} = b_j \frac{\bar{X}_j}{\bar{Y}} \quad (3.16)$$

и *бета-коэффициентов  $\beta_j$* :

$$\beta_j = b_j \frac{S_{x_j}}{S_y} \quad (3.17)$$

где  $S_{x_j} = \sqrt{\frac{\sum (X_j - \bar{X})^2}{n - (p + 1)}}$ ,  $S_y = \sqrt{\frac{\sum (Y_j - \bar{Y})^2}{n - (p + 1)}}$  – несмещенные

среднеквадратичные отклонения факторов  $X_j$  и  $Y$ .

***Коэффициенты эластичности  $E_j$  показывают, на сколько процентов изменится значение переменной  $Y$  при изменении  $X_j$  на 1%.***

***Бета-коэффициенты  $\beta_j$  показывают, на какую часть среднеквадратичного отклонения изменится  $Y$  при изменении  $X_j$  на величину своего среднеквадратичного отклонения при фиксированных значениях других переменных.***

Долю влияния  $j$ -го фактора в суммарном влиянии всех факторов на зависимую переменную  $Y$  оценивают с помощью *дельта-коэффициентов*:

$$\Delta_j = r_{YX_j} \frac{\beta_j}{R^2} \quad (3.18)$$

где  $r_{YX_j}$  – коэффициент корреляции между  $j$ -м фактором и переменной  $Y$ ,

$R^2$  – коэффициент детерминации.

Ранжирование факторов можно выполнить и с помощью *частной корреляции*.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и данным фактором при устранении влияния других факторов, включенных в уравнение регрессии. То есть, они позволяют выявить «чистую» зависимость признака от одного из факторов и установить, каково было бы влияние этого фактора на величину признака при условии, что влияние других (другого) факторов на этот признак исключается.

Обозначим  $R^2_{yx_1 x_2 \dots x_i \dots x_p}$  - множественный

коэффициент детерминации всего комплекса  $p$  факторов с результатом;

$R^2_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_p}$  - тот же показатель

детерминации, но без введения в модель фактора  $X_i$ .

Тогда в общем виде при наличии  $p$  факторов **коэффициент частной корреляции**, измеряющий влияние на  $Y$  фактора  $X_j$  при неизменном уровне других факторов, можно определить по формуле:

$$r_{yx_i \cdot x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_p} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \dots x_i \dots x_p}}{1 - R^2_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_p}}}$$

Частный F-критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении множественной регрессии. В общем виде для фактора  $x_j$  частный F-критерий определится как

$$F_{\text{част}x_i} = \frac{R^2_{yx_1 \dots x_i \dots x_p} - R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_p}}{1 - R^2_{yx_1 \dots x_i \dots x_p}} \cdot \frac{n - p - 1}{1}$$

где  $p$  – число факторов в линейном уравнении регрессии;

$n$  – число наблюдений.

Фактическое значение частного F-критерия сравнивается с табличным при 5%-ном или 1%-ном уровне значимости и числе степеней свободы:  $v_1 = 1$ ,  $v_2 = n - p - 1$ . Если фактическое значение превышает табличное, то дополнительное

Частный F-критерий в числителе содержит прирост доли объясненной дисперсии показателя  $Y$  за счет дополнительного включения в модель соответствующего фактора; в знаменателе – долю остаточной дисперсии по полной регрессионной модели.

включение соответствующего фактора в модель статистически оправдано, в противном случае фактор в модель включать нецелесообразно.

## 3.2 Использование модуля *Анализ данных EXCEL*

### 3.2.1 Оценивание коэффициентов корреляции: *Анализ данных / Корреляция*

Режим *Корреляция* модуля *Анализ данных* осуществляет вычисление парных коэффициентов корреляции линейной множественной регрессии.

Порядок действий:

1) проверьте доступ к **Пакету анализа**. При его отсутствии в главном меню выберите **Файл/Параметры/Надстройки**. Установите флажок **Пакет анализа**. Щелкните по кнопке **ОК**;

2) в главном меню выберите **Данные/Анализ данных /Корреляция**. Щелкните по кнопке **ОК**; после вызова режима *Корреляция* на экране появляется диалоговое окно, в котором задаются следующие параметры (рис. 3.2).

	Y	X1	X2	X3	X4
Y	1				
X1	0,788502	1			
X2	-0,93241	-0,66582	1		
X3	0,871233	0,812454	-0,89865	1	
X4	0,900588	0,778557	-0,93105	0,934773	1

Рисунок 3.2 - Диалоговое окно режима *Корреляция*

Результатом применения модуля *Корреляция* является нижний треугольник матрицы парных коэффициентов корреляции (рис. 3.1), который представляет собой зеркальное отражение верхнего треугольника (в силу симметричности матрицы).

### 3.2.2 Оценивание параметров линейной регрессии: *Анализ данных / Регрессия*

Режим *Регрессия* модуля *Анализ данных* осуществляет вычисление коэффициентов линейной множественной регрессии с  $p$  переменными, построение доверительных интервалов и проверку значимости уравнения регрессии.

В главном меню выберите *Данные/Анализ данных /Регрессия*. Щелкните по кнопке **ОК**; после вызова режима *Регрессия* на экране появляется диалоговое окно (рис. 3.3), в котором задаются следующие параметры:

1. *Входной интервал Y* – вводится диапазон адресов ячеек, содержащих значения  $y_i$  (ячейки должны составлять один столбец).

	A	B	C	D	E	F	G	H	I	J	K	L
1	Произв.	Ф/вооружен.	текучесть	Потери	Стаж							
2	труда	труда	кадров	раб.врем.	работы							
3	52	72	13	2,7	5							
4	53	74	12,5	2,8	5,5							
5	50	72	12	3	5							
6	51	73	11	3,2	6							
7	54	70	10,1	3,2	7							
8	55	67	9	3,3	8							
9	57	67	8,5	3,4	10							
10	52	62	8,2	3,6	10							
11	60	72	8	3,7	10,5							
12	60	72	5,5	3,7	11							
13	62	74	5	3,4	13							
14	64	75	4,7	4	10							
15	65	76	4,6	4,2	12							
16	67	80	4	4,3	13							
17	67	82	4,1	4,7	14							
18	62	84	4,2	4,8	14,5							
19	63	84	4,5	4,8	15,5							
20	66	87	4	4,9	17							
21	68	90	4	5	16,5							
22	70	92	3	4,7	17,5							
23												
24												

**Регрессия**

**Входные данные**

Входной интервал Y: \$A\$3:\$A\$22

Входной интервал X: \$B\$3:\$E\$22

Метки  Константа - ноль

Уровень надежности: 99 %

**Параметры вывода**

Выходной интервал: \$A\$24

Новый рабочий лист:

Новая рабочая книга

**Остатки**

Остатки  График остатков

Стандартизованные остатки  График подбора

**Нормальная вероятность**

График нормальной вероятности

OK Отмена Справка

Рисунок 3.3 - Диалоговое окно режима *Регрессия*

2. *Входной интервал X* – вводится диапазон адресов ячеек, содержащих значения независимых переменных. Значения каждой переменной представляются одним столбцом.

3. *Метки* – включается, если первая строка во входном диапазоне содержит заголовок. В этом случае автоматически будут созданы стандартные названия.

4. *Уровень надежности* – при включении этого параметра задается надежность  $1-\alpha$  при построении доверительных интервалов.

5. *Константа-ноль* – при включении этого параметра коэффициент  $b_0 = 0$ .

6. *Выходной интервал* – при включении активизируется поле, в которое необходимо ввести адрес левой верхней ячейки выходного диапазона, который содержит ячейки с результатами вычислений режима *Регрессия*.

7. *Новый рабочий лист* – при включении этого параметра открывается новый лист, в который начиная с ячейки A1 вставляются результаты работы режима *Регрессия*.

8. *Новая рабочая книга* – при включении этого параметра открывается новая книга на первом листе которой начиная с ячейки A1 вставляются результаты работы режима *Регрессия*.

9. *Остатки* – при включении вычисляется столбец, содержащий остатки (невязки)  $y_i - \hat{y}_i, i = 1, \dots, n$ .

10. *Стандартизованные остатки* – при включении вычисляется столбец, содержащий стандартизованные остатки.

11. *График остатков* – при включении выводятся точечные графики невязки  $y_i - \hat{y}_i, i = 1, \dots, n$ , в зависимости от значений переменных  $x_j$ . Количество графиков равно числу  $p$  переменных  $x_j$ .

12. *График подбора* – при включении выводятся точечные графики предсказанных по построенной регрессии значений  $\hat{y}_i$  от значений переменных  $x_j$ . Количество графиков равно числу  $p$  переменных  $x_j$ .

Например, в столбцах В, С, D и E листа Excel введены по двадцать значений четырех независимых переменных, а в столбец А – значения зависимой переменной (рис. 3.3).

В диалоговом окне режима *Регрессия* зададим необходимые параметры (рис. 3.3). Результаты работы программы приводятся на рис. 3.4 и рис. 3.5.

	A	B	C	D	E	F	G	H	
23									
24	ВЫВОД ИТОГОВ								
25									
26	<b>Регрессионная статистика</b>								
27	Множественный R	0,964185741							
28	R-квадрат	0,929654143							
29	Нормированный R-квадрат	0,910895248							
30	Стандартная ошибка	1,917239037							
31	Наблюдения	20							
32									
33	<b>Дисперсионный анализ</b>								
34		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
35	Регрессия	4	728,6629171	182,1657293	49,55804325	1,80251E-08			
36	Остаток	15	55,1370829	3,675805526					
37	Итого	19	783,8						
38									
39		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 99,0%</i>	<i>Верх</i>
40	Y-пересечение	56,91243209	7,568860683	7,519286518	1,83139E-06	40,77978743	73,04507676	34,6091728	7
41	Переменная X 1	0,337496653	0,101896176	3,312162092	0,004739229	0,120310094	0,554683212	0,037237877	0
42	Переменная X 2	-1,840601102	0,402267501	-4,575565009	0,00036422	-2,698013985	-0,98318822	-3,02596793	-0
43	Переменная X 3	-2,272155592	1,899118933	-1,196426171	0,250099566	-6,320031778	1,775720594	-7,868313818	3
44	Переменная X 4	-0,097584166	0,392655308	-0,248523741	0,807099626	-0,934509145	0,739340813	-1,254626622	3
45									

Рисунок 3.4 - Результаты работы режима *Регрессия*

Рассмотрим показатели, величины которых вычисляются в режиме *Регрессия*.

**1) Регрессионная статистика** (рис. 3.4).

*Множественный R* - корень квадратный из коэффициента детерминации.

*R – квадрат* – коэффициент детерминации  $R^2$ .

*Нормированный R–квадрат* – скорректированный на число степеней свободы – коэффициент детерминации  $\overline{R^2}$  (см. формулу (2.9)).

*Стандартная ошибка* – оценка  $s$  для среднеквадратического отклонения  $\sigma$ .

*Наблюдения* – число наблюдений  $n$ .

**2) Дисперсионный анализ** (рис. 3.4).

*Столбец df* - число степеней свободы, которое связано с числом наблюдений  $n$  и с числом определяемых оценок. Для строки *Регрессия* этот показатель равен числу независимых переменных ( $k_r = p$ ); для строки *Остаток* –  $k_e = n - (p + 1)$ ; для строки *Итого* –  $k = n - 1$ . Для условий рассматриваемого

примера число степеней свободы: для регрессии  $p = 4$ ; для остатка  $n - (p + 1) = 20 - (4 + 1) = 15$ ; общее  $n - 1 = 20 - 1 = 19$ ;

*Столбец SS* – сумма квадратов отклонений:

для строки *Регрессия* показатель (регрессионная сумма квадратов) равен величине  $Q_r$ :

$$SS_R = Q_R = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = 728,6629171 \text{—регрессионная сумма квадратов (объясненная);}$$

для строки *Остаток* показатель (остаточная сумма квадратов) равен величине  $Q_e$ :

$$SS_e = Q_e = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 55,1370829 \text{—остаточная сумма квадратов (не объясненная);}$$

для строки *Итого* показатель (общая сумма квадратов) равен  $Q = Q_r + Q_e$

$$\sum_{i=1}^n (Y - \bar{Y})^2 = 783,8 \text{—общая сумма квадратов (зависимой переменной);}$$

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y - \bar{Y})^2 = 728,6629 + 55,13708 = 783,8.$$

*Столбец MS* – дисперсии, вычисленные по формуле  $MS = \frac{SS}{df}$ , т.е. дисперсия в

расчете на одну степень свободы.

$$S_p^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{p} = 182,1657 \text{—регрессионная часть дисперсии } Y \text{ (объясненная}$$

моделью);

$$S_e^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - (p + 1)} = 3,675806 \text{—остаточная часть дисперсии } Y \text{ (не объясненная}$$

моделью).

*Столбец F* – величина  $F$  – критерия Фишера, вычисленная по формуле:

$$F_c = \frac{\frac{SS_r}{p}}{\frac{SS_e}{n-(p+1)}} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 (n - p - 1)}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 p} = 49,558$$

*Столбец значимость F* - значение уровня значимости, соответствующее вычисленной величине  $F$  – критерия и равное вероятности  $P(F(k_r, k_e) \geq F_c)$ , где  $F(k_r, k_e)$  - случайная величина, подчиняющаяся распределению Фишера с  $k_r, k_e$  степенями свободы. Эту вероятность можно также определить с помощью функции  $F_{РАСП}(F_c; k_r; k_e)$ . Если вероятность меньше уровня значимости  $\alpha$  (обычно  $\alpha = 0.05$ ), то построенная регрессия является значимой.

Уровень значимости F-критерия  $\alpha = 1,8E-08 = 0,00000018$ . Коэффициент детерминации  $R^2$  равен 0,9296, что указывает на сильную зависимость между независимыми переменными и производительностью труда. Можно использовать F-статистику, чтобы определить, является ли этот результат (с таким высоким значением  $R^2$ ) случайным. Величина  $\alpha$  применяется для обозначения вероятности ошибочного вывода о том, что имеется сильная взаимозависимость.

Из таблицы дисперсионного анализа следует, что вероятность случайно получить такое значение F-критерия ( $F = 49,558$ ) составляет  $P = 0,00000018$ , что не превышает допустимый уровень значимости 5% (величину  $\alpha = 0,05$ ). Следовательно, полученное значение не случайно, оно сформировалось под влиянием существенных факторов, то есть, подтверждается статистическая значимость всей модели и коэффициента детерминации.

*Столбец Коэффициенты* – вычисленные значения коэффициентов  $b_0, b_1, b_2, \dots, b_p$ , расположенных сверху-вниз.

*Столбец Стандартная ошибка оценки* – значения  $s_{b_j}, j = 0, 1, \dots, p$ , вычисленные по формуле  $s_{b_j} = \sqrt{s^2 \cdot \left\{ (X^T X)^{-1} \right\}_{j,j}}$ .

*Столбец t-статистика* – значения статистик  $t_j$ .

*Столбец P – значение* – содержит вероятности случайных событий  $P(t(n-m) \geq T_{b_j})$ , где  $t(n-m)$  – случайная величина, подчиняющаяся распределению Стьюдента с  $n-m$  степенями свободы.

Если эта вероятность меньше уровня значимости  $\alpha$ , то принимается гипотеза о значимости соответствующего коэффициента регрессии.

Из рис. 3.4 видно, что значимым коэффициентом являются только коэффициенты  $b_0, b_1, b_2$ .

Столбцы Нижние 95% и Верхние 95% с вероятностью 0,95 определяют верхние и нижние границы параметров модели  $\beta_j$ , т.е. позволяют осуществить интервальное оценивание параметров.

Так, например, для  $\beta_1$ : для  $\alpha = 5\%$  и  $20 - 5 = 15$  степеней свободы табличное значение (двустороннее)  $t_\alpha = 2,13$ , тогда

$$S_{b_1} \cdot t_\alpha = 0,1019 \cdot 2,131 = 0,01335;$$

$$0,3375 - 0,01335 \leq \beta_0 \leq 0,3375 + 0,01335;$$

$$0,1204 \leq \beta_1 \leq 0,5546.$$

Таким образом, с вероятностью 0,95, увеличение фондовооруженности труда на одну единицу обеспечит прирост производительности труда не ниже 0,1204 и не выше 0,55464 тыс. руб./чел.-ч.

Рассмотрим группу показателей, объединенных в таблице, показанной на рис. 3.5.

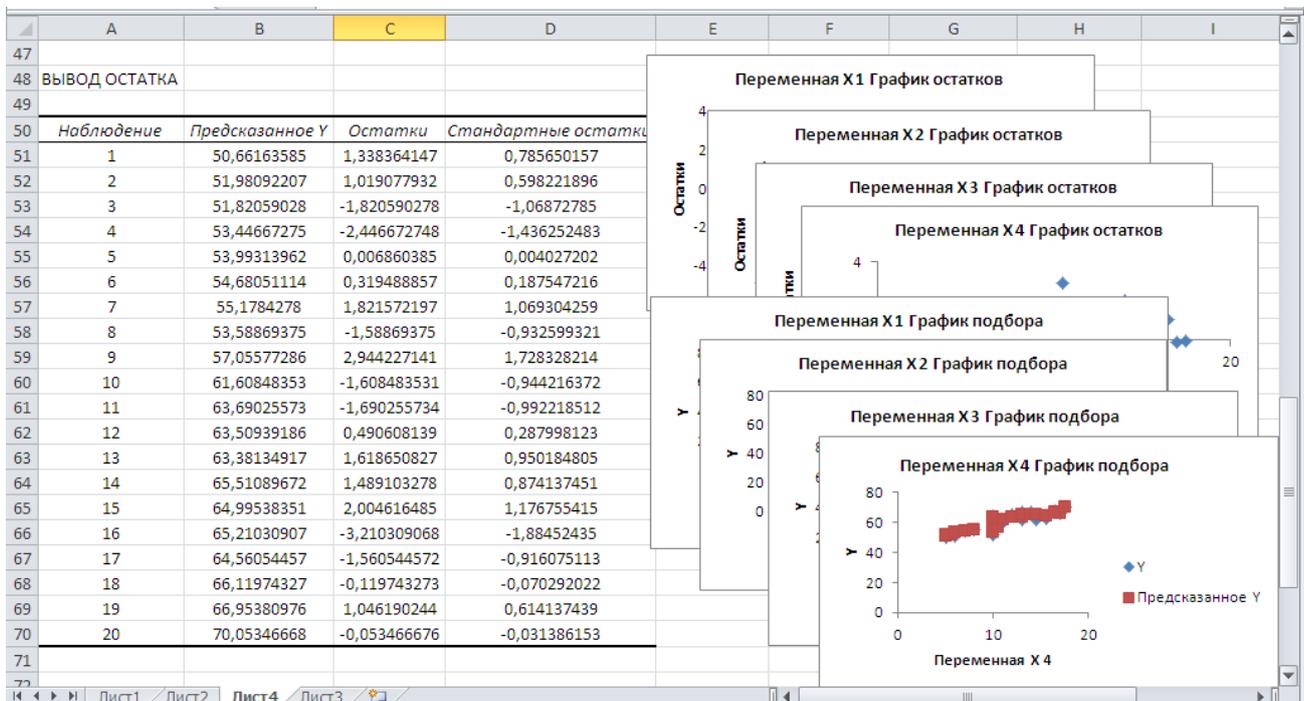


Рисунок 3.5 - Продолжение результатов работы режима Регрессия

*Столбец Наблюдение* – содержит номера наблюдений.

*Столбец Предсказанное Y* – значения  $\hat{y}_i$ , вычисленные по построенному уравнению регрессии.

*Столбец Остатки* – значения остатков  $y_i - \hat{y}_i$

Кроме того, режим *Регрессия* выводит графики остатков  $y_i - \hat{y}_i$  при заданных значениях независимых переменных. Изучение этих графиков позволяет составить первоначальное представление о гетероскедастичности и автокорреляции остатков

Так, наличие чередующихся положительных и отрицательных значений остатков является косвенным признаком *отсутствия систематической ошибки* (неучтенной независимой переменной) в построенном уравнении регрессии.

Программа *Анализ данных/Регрессия* позволяет вывести и графики «подбора», т.е. зависимости результативного признака от каждого из факторов, а также графики остатков для парных зависимостей. Для вывода графиков следует в окне **Регрессия** поставить соответствующие флажки.

### **3.3 Проверка качества регрессионной модели (коэффициент детерминации), информационные критерии**

Качество построения эконометрической модели связано с возможностью независимых переменных (регрессоров) объяснять изменчивость зависимой переменной. С этой целью используется коэффициент детерминации (табл. 3.1). Например, для модели, характеризующей изменчивость уровня заработной платы от образования человека и опыта работы значение  $R^2 = 0,31$  показывает, что только 31% изменений в зарплате объясняются изменениями в образовании человека и его опытом работы, остальные 69% , видимо, объясняются другими аспектами. Однако делать вывод о том, что модель не пригодна для анализа и бесполезна, тоже не стоит – результаты просто показывают относительно небольшой вклад образования и опыта работы у конкретного работодателя в объяснение заработной платы.

Таблица 3.1 – Коэффициент детерминации

Определение	Формула
Коэффициент детерминации ( $R^2$ ) показывает, какая доля дисперсии зависимой переменной объясняется регрессионным уравнением. При наличии константы в модели коэффициент детерминации изменяется от 0 до 1, причем, чем ближе к 1, тем лучше считается качество модели (то есть большая доля дисперсии зависимой переменной объясняется уравнением)	$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$ <p>RSS – сумма квадратов остатков, обусловленная регрессией;  ESS – сумма квадратов остатков, обусловленная ошибками;  TSS – общая сумма квадратов остатков регрессии</p>

Добавление регрессоров в модель будет способствовать росту коэффициента детерминации, даже если добавить регрессор с незначимым коэффициентом, который не улучшает объяснение дисперсии зависимой переменной.

Для сравнения качества регрессий с разным числом регрессоров используют скорректированный коэффициент детерминации:

$$R_{adj}^2 = 1 - \frac{ESS/(n - p - 1)}{TSS/p}$$

В главе 2 подробно рассматривался вопрос о проверке статистической надежности результатов оценивания. И для множественной регрессионной модели обязательно должна проводиться аналогичная проверка. В табл. 3.2 и 3.3. представлена методика такого исследования.

В начале проверим гипотезу о незначимости уравнения в целом с помощью теста Фишера (табл. 3.2).

Для проверки значимости коэффициентов регрессии при отдельных регрессорах используется тест Стьюдента (табл. 3.3).

Таблица 3.2 – Проверка статистической надежности уравнения регрессии

Тест Фишера позволяет проверить незначимость регрессии, то есть установить, равен ли нулю соответствующий параметр модели		
Гипотезы	Расчетная статистика	Правило принятия решения (гипотеза $H_0$ отклоняется, если):
$H_0 : \beta_1 = \dots = \beta_p = 0$ $H_1 : \text{как минимум один из коэффициентов отличен от нуля}$	$F_{rach} = \frac{RSS/p}{ESS/(n-p-1)}$ <p>RSS – сумма квадратов остатков, обусловленная регрессией;            ESS – сумма квадратов остатков, обусловленная ошибками</p>	$F_{rach} > F_{\alpha,(p; n-p-1)}$ <p>где <math>p</math> – число регрессоров в модели;  <math>\alpha</math> – уровень значимости;  <math>n</math> – число наблюдений.</p>

Таблица 3.3 – Проверка статистической надежности коэффициентов регрессии

Тест Стьюдента позволяет проверить незначимость отдельных коэффициентов регрессии, то есть, установить, равен ли нулю соответствующий параметр модели		
Гипотезы	Расчетная статистика	Правило принятия решения (гипотеза $H_0$ отклоняется, если):
$H_0 : \beta_j = 0$ $H_1 : \beta_j \neq 0$	$t_j = \frac{b_j}{s_j}$	$ t_j  > t_{\frac{\alpha}{2},(n-p)}$ <p>где <math>p</math> – число регрессоров в модели (включая свободный член);  <math>\alpha</math> – уровень значимости;  <math>n</math> – число наблюдений.</p>

Для сравнения моделей с целью выбора лучшей широкое распространение получили так называемые информационные критерии. Наиболее известным является критерий Акаике (AIC) (табл. 3.4).

Таблица 3.4 – Информационный критерий Акаике (AIC)

Определение	Формула
Информационный критерий Акаике (AIC) показывает степень качества модели и при этом вводит штрафную функцию за переизбыток регрессоров в модели. При прочих равных условиях следует предпочесть модель, в которой значение AIC более низкое.	$AIC = 2(p + 1) - 2\ln(L),$ <p>где <math>p</math> – число регрессоров в модели;  <math>L</math> – функция максимума правдоподобия для оцениваемой модели</p>

Второй популярный критерий, аналогичный критерию Акаике (AIC), это критерий Хеннана-Куинна (HQC) (табл. 3.5).

Таблица 3.5 – Информационный критерий Хеннана-Куинна (HQC)

Определение	Формула
Информационный критерий Хеннана-Куинна (HQC) – при прочих равных условиях следует предпочесть модель, в которой значение HQC более низкое.	$HQC = n \cdot \log\left(\frac{RSS}{TSS}\right) + 2(p + 1)\log(n),$ <p>где <math>p</math> – число регрессоров в модели;  RSS – сумма квадратов остатков, обусловленная регрессией;  <math>n</math> – число наблюдений.</p>

Еще один широко используемый информационный критерий – это Байесовский информационный критерий (BIC), он еще называется критерий Шварца (SC) (табл. 3.6).

Таблица 3.6 – Информационный критерий Шварца (SC)

Определение	Формула
Информационный критерий Шварца (SC) – при прочих равных условиях следует предпочесть модель, в которой значение SC более низкое.	$SC = (p + 1) \cdot \ln(n) - 2\ln(L),$ <p>где <math>p</math> – число регрессоров в модели;  <math>L</math> – функция максимума правдоподобия для оцениваемой модели;  <math>n</math> – число наблюдений.</p>

При выборе модели из нескольких имеющихся можно руководствоваться результатами всех трех критериев и выбирать ту модель, в которой хотя бы два из трех критериев имеют наименьшее значение.

### Вопросы для самоконтроля

1. Как определяется модель множественной линейной регрессии?
2. Перечислите предпосылки МНК. Каковы последствия их невыполнимости?
3. Что характеризуют коэффициенты регрессии?
4. Как определяется статистическая значимость коэффициентов регрессии?
5. Чем скорректированный коэффициент детерминации отличается от обычного?
6. Какие значения могут принимать обычный и скорректированный коэффициент детерминации при наличии свободного члена в уравнении регрессии?
7. Как осуществляется анализ статистической значимости коэффициента детерминации?
8. Как строятся интервальные оценки коэффициентов регрессии и в чем их суть?

## ТЕМА 4. МУЛЬТИКОЛЛИНЕАРНОСТЬ

### 4.1 Понятие мультиколлинearности

При изучении многофакторной линейной регрессии часто сталкиваются с наличием линейной связи между всеми или несколькими объясняющими переменными. Это явление называется *мультиколлинearностью*.

Как известно, одна из предпосылок метода наименьших квадратов постулирует отсутствие мультиколлинearности факторов, т.е. *объясняющие переменные не должны быть связаны тесной линейной зависимостью*. При нарушении данной предпосылки возникают следующие проблемы (рис. 4.1).

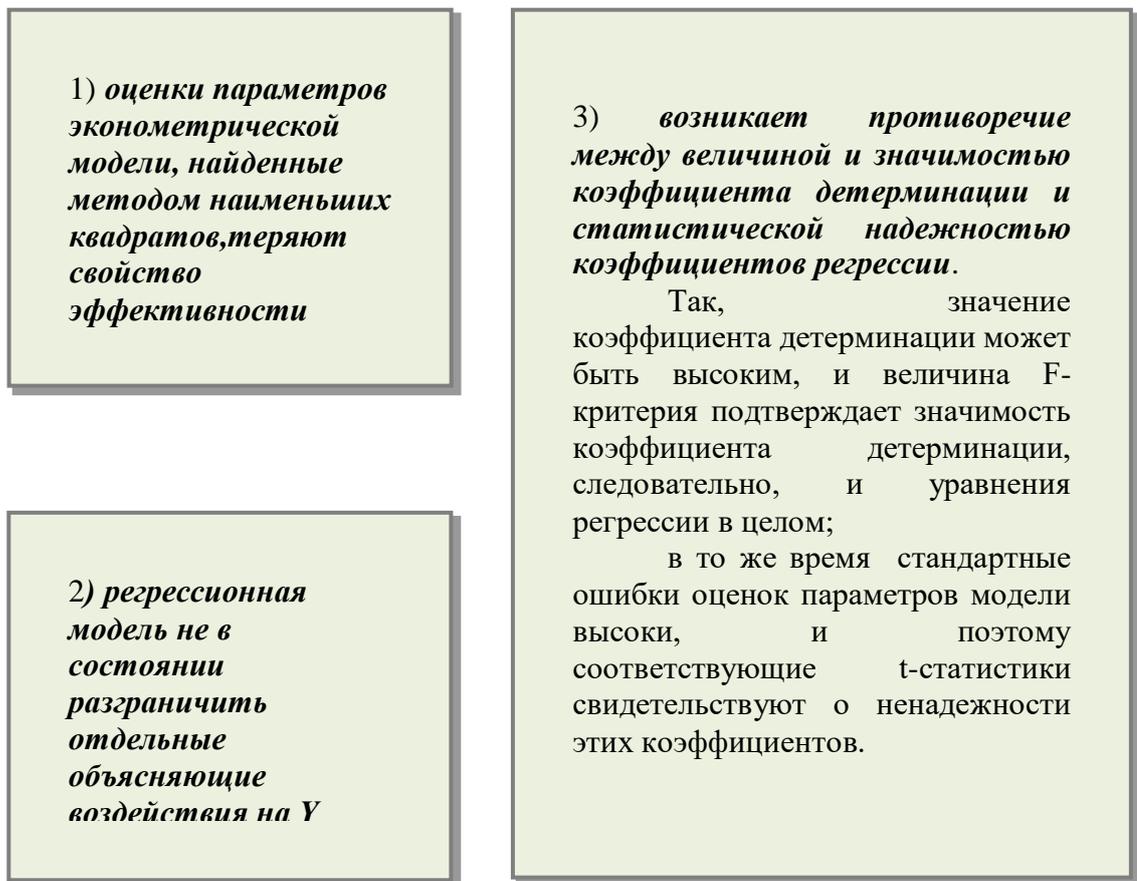


Рисунок 4.1 - Основные проблемы, возникающие при мультиколлинearности

**Признаки мультиколлинеарности:**

- наличие среди парных коэффициентов корреляции между факторами близких по величине множественному коэффициенту корреляции;
- если определитель матрицы парных коэффициентов корреляции между факторами ( $|R|$ ) равен нулю, то существует *полная* мультиколлинеарность, если же этот определитель равен единице, то мультиколлинеарность отсутствует; чем ближе к нулю величина определителя  $|R|$ , тем увереннее можно утверждать, что между независимыми переменными имеется мультиколлинеарность;
- в эконометрической модели при статистической надежности коэффициента детерминации имеются незначимые оценки параметров;
- наличие частных коэффициентов детерминации между объясняющими переменными, близких по величине к единице; сравнивая между собой эти коэффициенты можно найти переменные, ответственные за мультиколлинеарность. Для этого в качестве зависимой переменной рассматривается каждый из факторов. Чем ближе значение коэффициента детерминации к единице, тем сильнее проявляется мультиколлинеарность факторов.
- если при построении эконометрической модели на основе пошаговой регрессии включение новой переменной в модель существенно изменяет оценки параметров модели при незначительном повышении (или понижении) коэффициентов корреляции или детерминации, то, очевидно, эта переменная находится в линейной зависимости от остальных, которые были введены в уравнение ранее.

Все эти признаки имеют один недостаток: ни один из них четко не разграничивает случаи, когда мультиколлинеарность следует считать «существенной» и непременно учитывать, и когда мультиколлинеарностью можно пренебречь.

## 4.2 Исследование мультиколлинеарности

Наиболее полное исследование мультиколлинеарности можно осуществить с помощью **критерия Фаррара-Глоубера**, алгоритм которого реализуется в три этапа (рис. 4.2).

**На первом этапе** мультиколлинеарность выявляется в общем виде

Построение корреляционной матрицы объясняющих переменных  $R$  с помощью программы Анализ данных/Корреляция

Вычисление определителя матрицы  $R$  с помощью функции МОПРЕД (Excel,  $f_x$ , категория Математические).

Вычисление критерия  $\chi^2$ :

$$\chi^2 = -\left[ n-1 - \frac{1}{6}(2p+5) \right] \ln|R|$$

где  $|R|$  - определитель матрицы  $R$ ,  $n$  – число наблюдений  $i = \overline{1, n}$ ;  $p$  – количество объясняющих переменных  $j = \overline{1, p}$

Сравнение  $\chi^2$  с табличным при  $\nu = \frac{1}{2}p(p-1)$  степенях свободы и уровне значимости  $\alpha$ . Если  $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$ , то гипотеза  $H_0$  об отсутствии мультиколлинеарности отклоняется, то есть, в массиве независимых переменных существует мультиколлинеарность.

**На втором этапе** определяются переменные, тесно связанные с остальными

Вычисление матрицы, обратной для матрицы корреляций с помощью функции МОБР (Excel,  $f_x$ , категория Математические):  $C = R^{-1}$

Расчет F-критерия для каждой объясняющей переменной:

$$F_j = \left( c_{jj} - 1 \right) \frac{n-p}{p-1}$$

где  $c_{jj}$  - диагональные элементы матрицы  $C$

Сравнение  $F_j$  с табличным при  $\nu_1 = (p-1)$  и  $\nu_2 = (n-p)$  степенях свободы и уровне значимости  $\alpha$ . Если  $F_{j\text{факт}} > F_{j\text{табл}}$ , то соответствующей  $j$ -й переменной в наибольшей степени свойственна мультиколлинеарность.

На третьем этапе выявляются пары взаимосвязанных переменных

Расчет частных коэффициентов корреляции для каждой пары переменных:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} c_{jj}}}$$

где  $c_{kj}$  - элемент матрицы С, находящийся в  $k$ -й строке и  $j$ -м столбце;  $c_{kk}$  и  $c_{jj}$  - диагональные элементы матрицы С, ( $k, j = \overline{1, p}$ ).

Вычисление t-критериев для каждой пары переменных:

$$t_{kj} = \frac{r_{kj} \sqrt{n-p}}{\sqrt{1-r_{kj}^2}}$$

Фактические значения t-критериев сравнивают с табличными при  $n - p$  степенях свободы и уровне значимости  $\alpha$ . Если  $t_{kj \text{ факт}} > t_{\text{табл}}$ , то между объясняющими переменными  $X_k$  и  $X_j$  существует коллинеарность. Если  $t_{kj \text{ факт}} \leq t_{\text{табл}}$ , то данные не подтверждают наличие коллинеарности между переменными  $X_k$  и  $X_j$ .

Рисунок 4.2 - Исследование мультиколлинеарности с помощью критерия Фарара-Глоубера

### 4.3 Устранение мультиколлинеарности

Существуют различные подходы для преодоления сильной межфакторной мультиколлинеарности.

1. Наиболее простым способом является **исключение из модели одной из переменных мультиколлинеарной пары**. Однако, как уже отмечалось, часто удаление из модели некоторого фактора противоречит логике экономических связей. В экономических исследованиях эти вопросы должны решаться, прежде всего, исходя из логико-профессиональных соображений.

2. Другой метод состоит в **преобразовании объясняющих переменных**, которое предполагает один из следующих вариантов:

➤ линейное преобразование переменных – замена мультиколлинеарных

переменных их линейной комбинацией. Например, следует построить уравнение регрессии в виде:  $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$ . Установлено, что переменные  $X_1$  и  $X_2$  высоко коррелированы. Анализ рассматриваемого процесса и результаты наблюдений позволяют постулировать дополнительно уравнение связи между объясняющими переменными  $X_1$  и  $X_2$ , а именно  $X_2^* = X_1 - X_2$ . Переменную  $X_2^*$  подставляем в уравнение регрессии и получаем:  $\hat{Y} = b_0^* + b_1^* X_1 + b_2^* X_2^*$ . В общем случае переменные  $X_1$  и  $X_2^*$  не сильно коррелированы. Таким образом достигается снижение или даже полное устранение мультиколлинеарности;

- стандартизация объясняющих переменных – переход к стандартизованным (нормализованным) переменным иногда позволяет решить проблему мультиколлинеарности;
- исключение тренда – основной тенденции, характеризующей изменение рассматриваемого показателя  $Y$  во времени. При построении регрессии по данным, полученным из временных рядов, рекомендуется исключить тренд, вычитая из фактических значений зависимой переменной  $Y$  ее расчетные значения, полученные по модели, в которой  $Y$  представляется как функция времени. Этим достигается соблюдение предпосылок регрессионного анализа – независимость наблюдений и уменьшение мультиколлинеарности.

3. **Пошаговая регрессия.** Процедура применения пошаговой регрессии начинается с построения простой регрессии. На первом шаге рассматривается лишь одна объясняющая переменная, имеющая с зависимой переменной  $Y$  наибольший коэффициент детерминации.

На втором шаге включается в регрессию новая объясняющая переменная, которая вместе с первоначально отобранной образует пару объясняющих переменных, имеющую с  $Y$  наиболее высокий скорректированный коэффициент детерминации.

На третьем шаге вводится объясняющая переменная, которая вместе с двумя первоначально отобранными образует тройку объясняющих переменных, имеющую с  $Y$  наиболее высокий скорректированный коэффициент детерминации и т.д.

Таким образом, в анализ последовательно включается по одной объясняющей переменной. На каждом шаге проверяется значимость коэффициентов регрессии и оценивается мультиколлинеарность переменных. Если оценка коэффициента регрессии получается незначимой, то переменная исключается и рассматривают другую объясняющую переменную. Если оценка коэффициента регрессии значима, а мультиколлинеарность отсутствует, то в анализ включают следующую переменную.

Строгих математических правил «останова» этой процедуры, т.е. выбора оптимального числа  $k$  объясняющих переменных, которые нужно включить в модель, нет. Однако в большинстве случаев помочь решить этот вопрос может следующее.

Обратимся к рис. 4.3.

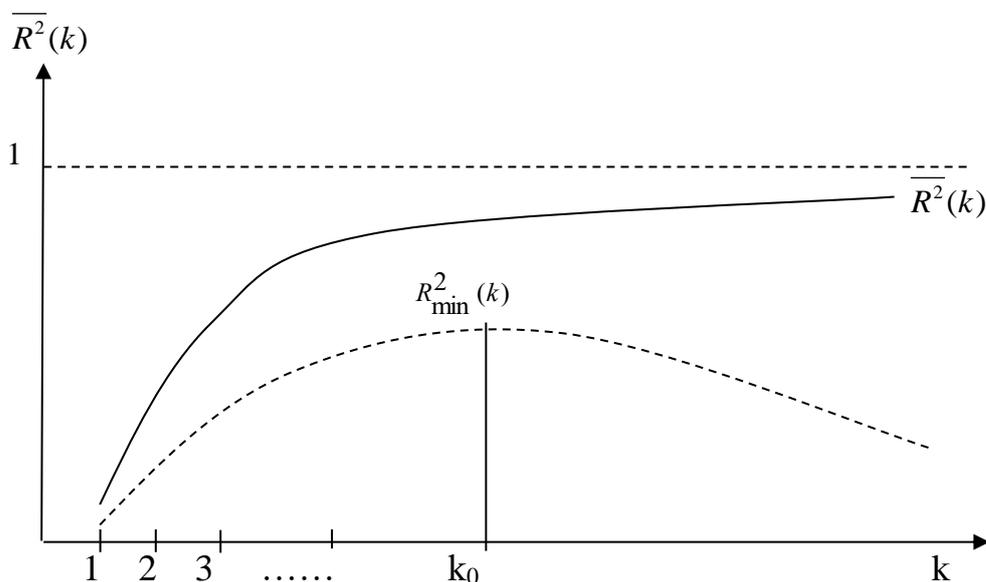


Рисунок 4.3 - Определение количества включаемых в модель переменных

По оси абсцисс (рис. 4.3) откладываем последовательно увеличивающееся на единицу число  $k$ -объясняющих переменных в модели, а по оси ординат –

соответствующие значения скорректированного на число степеней свободы (на несмещенность) коэффициента детерминации:

$$\overline{R^2} = 1 - (1 - R^2) \frac{n-1}{n-p-1}. \quad (4.1)$$

Одновременно нанесем на чертеж по оси ординат величину нижней доверительной границы  $R_{\min}^2$ , вычисленную по формуле (4.2):

$$R_{\min}^2 = \overline{R^2}(k) - 2 \cdot \sqrt{\frac{2k(n-k-1)}{(n-1)(n^2-1)}} (1 - R^2(k)) \quad (4.2)$$

***В качестве оптимального числа объясняющих переменных  $k_0$  регрессионной модели выбирается такое значение  $k$ , при котором величина  $R_{\min}^2(k)$  достигает своего максимума.***

Таким образом, постепенно определяются все составляющие регрессии без нарушения предположения об отсутствии мультиколлинеарности.

4. ***Гребневая регрессия (ридждж-регрессия)*** – метод, в соответствии с которым осуществляется переход от несмещенных оценок, определенных по МНК, к смещенным оценкам, обладающим, однако, меньшим рассеянием относительно оцениваемого параметра, т.е. меньшей дисперсией. Известно, что оценки, определяемые вектором

$$B = (X'X)^{-1} X'Y,$$

обладают минимальными дисперсиями в классе всех линейных *несмещенных* оценок, но при наличии мультиколлинеарности эти дисперсии могут оказаться слишком большими, и обращение к соответствующим *смещенным* оценкам может повысить точность оценивания параметров регрессии.

При использовании гребневой регрессии вместо несмещенных оценок рассматривают смещенные оценки, задаваемые вектором

$$B_\lambda = (X'X + \lambda I_{p+1})^{-1} X'Y,$$

где  $\lambda$  - некоторое положительное число, называемое «гребнем» или «хребтом»,  
 $I_{p+1}$  – единичная матрица  $(p+1)$ -го порядка,  
 $p$  – количество независимых переменных регрессии.

Стандартная ошибка  $S_{bk}$  ридж-оценки  $k$ -го коэффициента регрессии, равная произведению корня квадратного из соответствующего диагонального элемента ковариационной матрицы векторной оценки на стандартную ошибку остатков, представляется в виде:

$$S_{bk} = S_{\hat{Y}} \sqrt{(X'X + \lambda I_{p+1})^{-1}},$$

где

$$S_e = \frac{e'e}{n-p-1} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-p-1}}, \lambda = \frac{pS_e^2}{B'B}.$$

Добавление  $\lambda$  к диагональным элементам матрицы  $(X'X)$  делает оценки параметров модели смещенными, но при этом увеличивается определитель матрицы системы нормальных уравнений: вместо  $|X'X|$  он будет равен  $|X'X + \lambda E_{p+1}|$ . Таким образом, становится возможным исключение мультиколлинеарности в случае, когда определитель  $|X'X|$  близок к нулю, т.е. **при наличии значительных корреляций входных переменных** (когда оценки коэффициентов либо меняют знак, либо изменяется значимость коэффициентов). Гребневые оценки параметров регрессии хоть и смещены, но имеют лучшие характеристики точности.

Доказано, что существует такое  $\lambda$ , при котором ридж-оценки более эффективны, чем оценки МНК (МНК-оценки наиболее эффективны среди несмещенных оценок). Проблема использования метода гребневой регрессии на практике сводится к выбору подходящего значения  $\lambda$ . Вместо использования формулы (...) можно осуществить перебор значений в интервале от 0,1 до 0,4 (поскольку четких правил выбора этого параметра нет) с целью определения его оптимальной величины.

5. **Метод главных компонент.** Основная идея заключается в сокращении числа объясняющих переменных до наиболее влияющих факторов. Это достигается путем линейного преобразования всех объясняющих переменных в новые переменные, так называемые главные компоненты. При этом требуется, чтобы выделению первой главной компоненты соответствовал максимум общей дисперсии всех объясняющих переменных, второй компоненте – максимум оставшейся дисперсии, после того, как влияние первой главной компоненты исключается, и т.д. Таким образом, выполненное преобразование содействует уменьшению мультиколлинеарности новых выделенных переменных по сравнению с мультиколлинеарностью набора исходных переменных.

### **Вопросы для самоконтроля**

1. Объясните значение терминов «коллинеарность» и «мультиколлинеарность».
2. Каковы основные последствия мультиколлинеарности с точки зрения качества модели?
3. Как можно обнаружить мультиколлинеарность?
4. Назовите основные методы устранения мультиколлинеарности?
5. В чем суть метода пошаговой регрессии применительно к проблеме устранения мультиколлинеарности?

## ТЕМА 5. ИСПОЛЬЗОВАНИЕ ФИКТИВНЫХ ПЕРЕМЕННЫХ В ЭКОНОМЕТРИЧЕСКОМ АНАЛИЗЕ

### 5.1 Введение фиктивных переменных в линейную модель регрессии

До сих пор в качестве факторов рассматривались экономические переменные, принимающие количественные значения в некотором интервале. Экономические величины складываются под влиянием множества различных факторов, как количественных, так и качественных, поэтому при построении регрессионной модели часто возникает необходимость включить в модель фактор, имеющий два или более качественных уровней (например, должность, образование, пол, место проживания, национальность и др.) и исследовать их влияние на результирующую переменную.

Учет влияния на структуру модели *качественных факторов (сопутствующих переменных)* осуществляется, как правило, с помощью аддитивного введения в правую часть регрессионного уравнения определенного числа *фиктивных (дихотомических, бинарных, булевых)* переменных, т.е. таких переменных, которые могут принимать одно из двух возможных значений (ноль или единица) и тем самым отражают два противоположных состояния качественного фактора: фактор действует – фактор не действует (курс валюты фиксированный – курс валюты плавающий; сезон летний – сезон зимний и т.д.). Обычно это записывается так:

$$Z = \begin{cases} 1, & \text{фактор действует,} \\ 0, & \text{фактор не действует} \end{cases}$$

Конкретная форма введения фиктивных экзогенных переменных в модель (в соответствии с допущением о характере влияния качественных факторов на зависимую переменную) имеет различный вид.

1. Автономное влияние качественного признака. Такие модели, содержащие лишь качественные объясняющие переменные, называются ANOVA-моделями (моделями дисперсионного анализа).

Например,  $Y$  – начальный размер заработной платы, который зависит от уровня образования. С помощью фиктивной переменной такую зависимость можно представить в виде:

$$D = \begin{cases} 1, & \text{претендент на рабочее место имеет высшее образование,} \\ 0, & \text{претендент не имеет высшего образования} \end{cases}$$

Тогда модель можно записать так:

$$Y = \beta_0 + \beta_1 D + \varepsilon.$$

Если включаемый в рассмотрение качественный признак имеет не два, а несколько значений, то используют несколько фиктивных переменных. При этом следует соблюдать правило: **количество вводимых фиктивных переменных должно быть на единицу меньше, чем число градаций (возможных значений) признака**. В противном случае попадают в так называемую «ловушку фиктивных переменных», связанную с возникновением полной мультиколлинеарности.

Добавим в рассматриваемый пример условие, что признак «уровень образования» имеет три градации: высшее, среднее, начальное. Тогда следует ввести две фиктивные переменные:

$$D_1 = \begin{cases} 1, & \text{претендент на рабочее место имеет высшее образование,} \\ 0, & \text{для других уровней образования} \end{cases}$$

$$D_2 = \begin{cases} 1, & \text{претендент на рабочее место имеет среднее образование,} \\ 0, & \text{для других уровней образования} \end{cases}$$

Тогда модель примет вид:

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \varepsilon.$$

Выполнив оценку параметров методом наименьших квадратов, получим:

$$\hat{Y} = b_0 + b_1 D_1 + b_2 D_2.$$

Эта модель для разных претендентов позволяет прогнозировать размер заработной платы:

$$\begin{aligned} \text{Если образование высшее} \quad \hat{Y} &= \beta_0 + \beta_1 \cdot 1 + \beta_2 \cdot 0 = \beta_0 + \beta_1; \\ \text{среднее} \quad \hat{Y} &= \beta_0 + \beta_1 \cdot 0 + \beta_2 \cdot 1 = \beta_0 + \beta_2; \\ \text{начальное} \quad \hat{Y} &= \beta_0 + \beta_1 \cdot 0 + \beta_2 \cdot 0 = \beta_0 \end{aligned}$$

Как видно,  $\beta_0$  означает некоторую начальную заработную плату,  $\beta_1$  и  $\beta_2$  показывают, на какую величину изменится (надо полагать, повысится) заработная плата с учетом изменения уровня образования.

2. Больше применение находят модели, содержащие как качественные, так и количественные переменные, которые называются ANCOVA-моделями (моделями ковариационного анализа).

Например, при определении зависимости цены двухкомнатной квартиры от ее полезной площади ( $X$ ) учитывается тип дома: «хрущевка», панельный, кирпичный с помощью фиктивных переменных.

Поскольку используются три категории домов, то вводятся 2 фиктивные переменные:  $Z^{(1)}$  и  $Z^{(2)}$ . Пусть  $Z^{(1)}$  принимает значение 1 для панельного дома и 0 для всех остальных типов домов; переменная  $Z^{(2)}$  принимает значение 1 для кирпичных домов и 0 для остальных типов; тогда переменные  $Z^{(1)}$  и  $Z^{(2)}$  принимают значения 0 для домов типа «хрущевки».

Предположим, что уравнения регрессии с фиктивными переменными выглядит так:

$$\hat{Y} = 320 + 500X + 2200Z_1 + 1600Z_2.$$

Частные уравнения регрессии для отдельных типов домов, свидетельствуя о наиболее высоких ценах на квартиры в панельных домах, будут иметь вид:

- «хрущевки»  $\hat{Y} = 320 + 500X$ , ( $Z^{(1)} = Z^{(2)} = 0$ );
- панельные  $\hat{Y} = 2520 + 500X$ , ( $Z^{(1)} = 1, Z^{(2)} = 0$ ); здесь  $2520 = 320 + 2200$ ),
- кирпичные  $\hat{Y} = 1920 + 500X$ , ( $Z^{(1)} = 0, Z^{(2)} = 1$ ); (здесь  $1920 = 320 + 1600$ ).

Коэффициенты при фиктивных переменных  $Z^{(1)}$  и  $Z^{(2)}$  представляют собой разность между средним уровнем результативного признака для соответствующей группы и базовой группы.

В рассматриваемом примере за базу сравнения цены взяты дома «хрущевки», для которых  $Z^{(1)} = Z^{(2)} = 0$ .

Коэффициент при  $Z^{(1)} = 2200$  означает, что при одной и той же полезной площади квартиры цена ее в панельных домах в среднем на 2200 долл. США выше, чем в «хрущевках».

Соответственно коэффициент при  $Z^{(2)} = 1600$  показывает, что в кирпичных домах цена выше в среднем на 1600 долл. США при неизменной величине полезной площади по сравнению с указанным типом домов.

Рассмотрим *использование фиктивных переменных в сезонном анализе*. Многие экономические показатели напрямую связаны с сезонными колебаниями. Например, спрос на туристические путевки, на охлажденную воду и мороженное существенно выше летом, чем зимой. Спрос на обогреватели, шубы выше зимой.

Обычно сезонные колебания характерны для временных рядов. Исследователи, использующие данные временных рядов, обычно предпочитают по годовым данным сведения по кварталам, т.к. за счет этого они получают в 4 раза больше наблюдений в рассматриваемый период. Но при этом часто наблюдается заметное воздействие сезонного фактора. Если это воздействие не учитывать, то оно вносит свой вклад в случайную составляющую модели, вследствие чего происходит снижение эффективности оценок параметров. Устранение или нейтрализация сезонного фактора в таких моделях позволяет сосредоточиться на других важных количественных и качественных характеристиках модели, в частности, на общем направлении развития, так называемом *тренде*. Такое устранение сезонного фактора называется *сезонной корректировкой*.

Одним из методов сезонной корректировки (оценки *сезонных различий*) является *метод фиктивных переменных*.

Пусть переменная  $Y$  определяется количественной переменной  $X$ , причем эта зависимость существенно различается по кварталам. Тогда общую модель такой ситуации можно представить в виде:

$$Y_t = \beta_0 + \beta_1 X_t + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \varepsilon_t, \quad (5.1)$$

где  $Z_j$  – фиктивные переменные, введенные для описания сезонных различий:

$$Z_1 = \begin{cases} 1 - \text{для первого квартала} \\ 0 - \text{для остальных кварталов} \end{cases}$$

$$Z_2 = \begin{cases} 1 - \text{для второго квартала} \\ 0 - \text{для остальных кварталов} \end{cases}$$

$$Z_3 = \begin{cases} 1 - \text{для третьего квартала} \\ 0 - \text{для остальных кварталов} \end{cases}$$

В рассматриваемом случае четвертый квартал выступает в качестве базы для сравнения сезонных изменений. Если значения  $Y_t$  существенно различаются по кварталам (сезонам), то в уравнении (5.1) коэффициенты при фиктивных переменных окажутся статистически значимыми. Тогда ожидаемое значение  $Y$  по кварталам определится следующими соотношениями:

$$E(Y|Z_1 = 1, Z_2 = 0, Z_3 = 0) = (\beta_0 + \alpha_1) + \beta_1 X - \text{для первого квартала}$$

$$E(Y|Z_1 = 0, Z_2 = 1, Z_3 = 0) = (\beta_0 + \alpha_2) + \beta_1 X - \text{для второго квартала}$$

$$E(Y|Z_1 = 0, Z_2 = 0, Z_3 = 1) = (\beta_0 + \alpha_3) + \beta_1 X - \text{для 3-го квартала}$$

$$E(Y|Z_1 = 0, Z_2 = 0, Z_3 = 0) = \beta_0 + \beta_1 X - \text{для четвертого квартала}$$

Видим, что в модели (5.1) рассматриваются такие ситуации, при которых квартальные различия отражаются лишь в различии свободных членов моделей.

3. Если необходимо учесть структурные изменения модели, то используют **фиктивные переменные наклона**. При этом говорят о влиянии в виде эффекта взаимодействия сопутствующих переменных. То есть, изменение качественного фактора может привести к изменению как свободного члена уравнения, так и наклона прямой регрессии. Обычно это характерно для временных рядов экономических данных (в частности, при введении новых правовых или налоговых ограничений).

Для учета структурных изменений в уравнении регрессии фиктивную переменную вводят как множитель при количественной переменной  $x$  и тогда зависимость может быть выражена так:

$$Y = \beta_0 + \beta_1 x + \alpha_1 D + \alpha_2 D \cdot x + \varepsilon, \quad (5.2)$$

где

$$D = \begin{cases} 1, & \text{до изменения условий,} \\ 0, & \text{после изменения условий.} \end{cases}$$

В этом случае ожидаемое значение зависимой переменной определяется следующим образом:

$$\begin{aligned} \hat{Y} &= \beta_0 + \beta_1 x, & D &= 0, \\ \hat{Y} &= (\beta_0 + \alpha_1) + (\beta_1 + \alpha_2)x, & D &= 1. \end{aligned}$$

Коэффициенты  $\alpha_1$  и  $\alpha_2$  в уравнении (5.2) называются соответственно *дифференциальным свободным членом* и *дифференциальным угловым коэффициентом*. Фиктивная переменная  $D$  в уравнении (5.2) используется как в *аддитивном виде* ( $\alpha_1 D$ ), так и в *мультипликативном* ( $\alpha_2 D \cdot X_t$ ), что позволяет фактически разбивать рассматриваемую зависимость на две части, связанные, с периодами изменения некоторого рассматриваемого в модели качественного фактора. Поэтому и говорят о **структурных изменениях модели**.

Вернемся к анализу влияния сезонности. Если, в отличие от рассмотренного примера, различия затрагивают и изменение коэффициента регрессии при объясняющей переменной  $X$ , то это может быть учтено в следующей модели:

$$Y_t = \beta_0 + \beta_1 X_t + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_4 Z_1 X_t + \alpha_5 Z_2 X_t + \alpha_6 Z_3 X_t + \varepsilon_t, \quad (5.3)$$

Здесь  $\alpha_1, \alpha_2, \alpha_3$  – дифференциальные свободные члены,  $\alpha_4, \alpha_5, \alpha_6$  – дифференциальные угловые коэффициенты.

Выбор правильной формы модели регрессии является в данной ситуации достаточно серьезной проблемой, так как вполне вероятны ошибки спецификации. Поэтому рекомендуется следующая схема выбора модели.

Вначале строится модель (5.3) и оценивается статистическая значимость коэффициентов  $\alpha_j$ . Если дифференциальные угловые коэффициенты оказываются незначимыми, то переходят к модели (5.1). Если и в этой модели дифференциальные свободные члены оказываются статистически незначимыми, то делают вывод, что квартальные (сезонные) изменения несут существенны для рассматриваемой зависимости.

## 5.2 Проверка регрессионной однородности двух групп наблюдений

Итак, фиктивные переменные вводятся в эконометрическую модель с целью учета воздействия качественных аспектов на закономерности развития рассматриваемых процессов. На рис. 5.1 показано, что в период  $(0, t)$  для развития процесса была характерна одна тенденция, а в период  $(t, T)$  – тенденция изменилась. На рис. 5.1 (а) зависимость моделируется обычной линейной регрессией. На рис. 5.1 (б) в модели учитываются изменения, произошедшие с некоторого момента  $t$  в характере расположения точек наблюдений. Следует проанализировать структурные изменения и сделать вывод, имеет ли смысл разбиение выборки на части и построение для каждой из них уравнения регрессии (т.е. фактически следует ли строить сложную регрессию с фиктивными переменными) (рис. 5.1 (б)) или можно ограничиться общей регрессией для всех точек наблюдений (рис. 5.1 (а)).

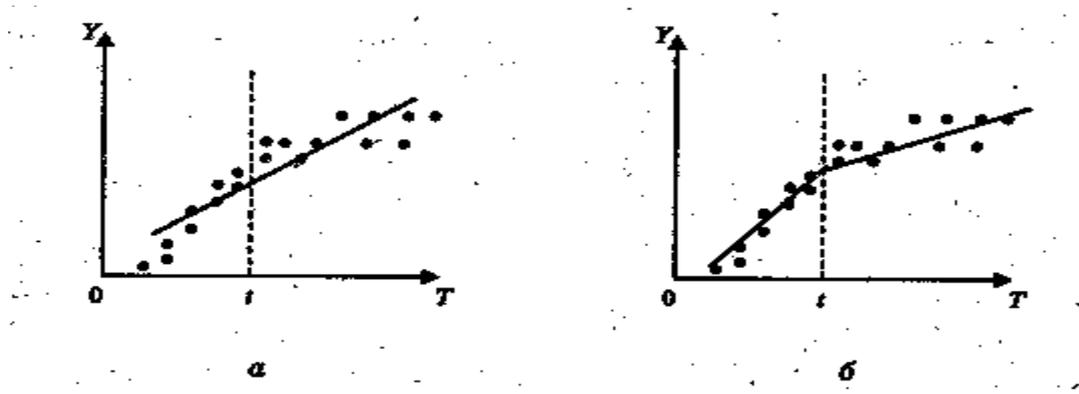


Рисунок 5.1 - Структурные изменения

Этот анализ предполагает получение ответа на вопрос: действительно ли подвыборки 1 и 2 неоднородны в регрессионном смысле или их можно объединить и строить искомую функцию регрессии по объединенной (общей) выборке. Другими словами, следует статистически проверить гипотезу

$$H_0: \beta^{(1)} = \beta^{(2)}, \quad D(\varepsilon_1) = D(\varepsilon_2) = \sigma^2,$$

где  $\beta^{(j)} = (\beta_0^{(j)}, \beta_1^{(j)}, \dots, \beta_p^{(j)})$  и  $\varepsilon_j$  – соответственно коэффициенты регрессии и случайные регрессионные остатки в классической линейной регрессионной модели, ( $j = 1, 2$ ).

Г. Чоу предложил следующую процедуру для статистической проверки гипотезы  $H_0$  в данном случае.

1) Выборка размера  $n$  разбивается на две подвыборки размерами  $n_1$  и  $n_2$  соответственно ( $n = n_1 + n_2$ ) и для каждой подвыборки строится уравнение регрессии, вычисляются векторы остатков ( $e_t = Y_t - \hat{Y}_t$ ) и суммы квадратов этих остатков  $S_1$  и  $S_2$ .

2) По объединенной (общей) выборке объем  $n$  определяются МНК-оценки параметров линейной модели (общего уравнения регрессии (рис. 5.1, а)), вычисляется вектор остатков  $e_t = Y_t - \hat{Y}_t$ , а затем сумма квадратов этих остатков:  $S_o = \sum e_t^2$ .

3) Критическая статистика определяется по формуле:

$$F = \frac{\frac{S_0 - (S_1 + S_2)}{p + 1}}{\frac{S_1 + S_2}{n_1 + n_2 - 2p - 2}},$$

или

$$F = \frac{S_0 - (S_1 + S_2)}{S_1 + S_2} \cdot \frac{n - 2p - 2}{p + 1}. \quad (5.4)$$

При этом если  $F < F_{\alpha; p+1; n-2p-2}$ , то считается, что различие между  $S_0$  и  $(S_1 + S_2)$  статистически незначимо и нет смысла разбивать уравнение регрессии на части. В противном случае разбиение на подинтервалы целесообразно с точки зрения улучшения качества модели. Это фактически означает необходимость введения в уравнение регрессии соответствующей фиктивной переменной.

Отметим, что использование теста Чоу осуществляется достаточно просто. Однако оно менее информативно, нежели общий анализ сложной регрессии с фиктивными переменными, осуществляемый на базе t-статистик (с учетом вклада каждой фиктивной переменной), коэффициента детерминации и статистики Дарбина – Уотсона. Однако тест Чоу вполне достаточен, если требуется установить, что зависимости в подвыборках различаются.

### Вопросы для самоконтроля:

1. В чем преимущества фиктивных переменных?
2. В чем суть ANOVA-моделей?
3. В чем суть ANCOVA-моделей?
4. В чем состоит правило применения фиктивных переменных?
5. Какой смысл имеет дифференциальный свободный член?
6. Какой смысл имеет дифференциальный угловой коэффициент?
7. Какова идея теста Чоу?
8. Как сезонные переменные применяются для устранения сезонного фактора?

## ТЕМА 6. РЕГРЕССИОННЫЕ МОДЕЛИ С ГЕТЕРОСКЕДАСТИЧНЫМИ ОСТАТКАМИ. ТЕСТИРОВАНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

### 6.1 Понятие гетероскедастичности

Прежде чем использовать построенную эконометрическую модель, важно определить, выполнены ли основные предпосылки МНК, поскольку от этого зависит, обладают ли оценки параметров модели нужными свойствами. Одной из таких предпосылок является постоянство дисперсии возмущений (гомоскедастичность).

Речь здесь идет о том, что при использовании некоторой совокупности наблюдений исследователь имеет дело с конкретными реализациями зависимой переменной  $Y_i$  и соответственно с определенными случайными отклонениями  $\varepsilon_i$ ,  $i = 1, 2, \dots, n$ . Но до осуществления выборки эти показатели априори могли принимать произвольные значения на основе некоторых вероятностных распределений. **Одним из требований к этим распределениям является равенство дисперсий.** Данное условие подразумевает, что, несмотря на то, что при каждом конкретном наблюдении случайное отклонение может быть большим либо маленьким, положительным либо отрицательным, не должно быть априорной причины, вызывающей большую ошибку (отклонение) при одних наблюдениях и меньшую – при других.

На рис. 6.1 приведены два примера линейной регрессии — зависимости потребления  $C$  от дохода  $I$  [3]:

$$C = \beta_0 + \beta_1 I + \varepsilon.$$

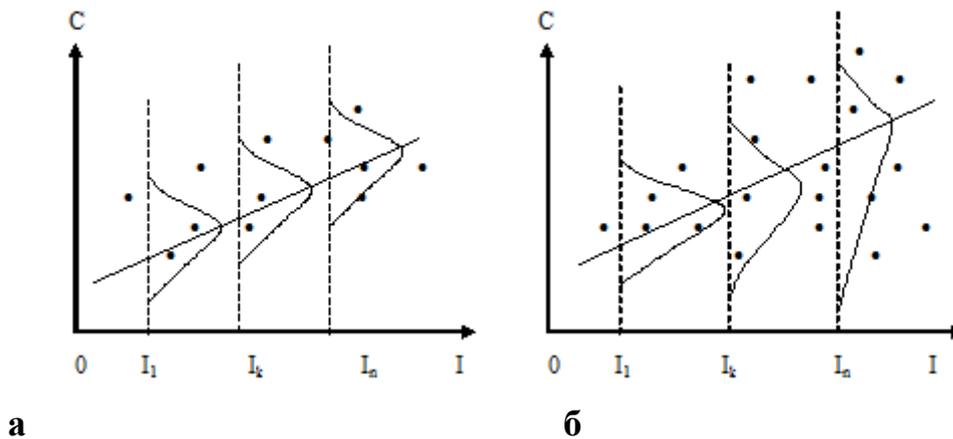


Рисунок 6.1- Изменение потребления в зависимости от изменения дохода

В обоих случаях с ростом дохода растет среднее значение потребления. Но если на рис. 6,а дисперсия потребления остается одной и той же для различных уровней дохода, то на рис. 6,б при аналогичной зависимости среднего потребления от дохода дисперсия потребления не остается постоянной, а увеличивается с ростом дохода. Фактически это означает, что во втором случае субъекты с большим доходом в среднем потребляют больше, чем субъекты с меньшим доходом, и, кроме того, разброс в их потреблении более существенен для большего уровня дохода.

Разброс значений потребления вызывает разброс точек наблюдения относительно линии регрессии, что и определяет дисперсию случайных отклонений. При гомоскедастичности (рис. 6, а) дисперсии  $\varepsilon_i$  постоянны, а при гетероскедастичности (рис. 6, б) дисперсии  $\varepsilon_i$  изменяются (в нашем примере увеличиваются).

На практике эта предпосылка часто не выполняется, так как дисперсии возмущений не является постоянной (меняется при переходе от наблюдения к наблюдению), то есть имеет место *гетероскедастичность*. Последствия гетероскедастичности представлены на рис. 6.2.

Невозможность применения t-статистик (рис. 6.2) можно пояснить следующим образом. Чаще всего смещение стандартных отклонений оценок параметров модели является отрицательным (т.е. в сторону уменьшения), значит,

стандартные ошибки будут меньше, чем они должны быть, а t-критерий – больше, чем в реальности. Из-за этого можно отвергнуть нулевую гипотезу, в то время как она должна быть принята, т.е. ошибочно будем считать коэффициент регрессии значимым, тогда как это не так.

Оценки параметров модели больше не представляют собой лучшие оценки, или, другими словами, не являются оценками с минимальной дисперсией, т.е. они не обладают свойством эффективности. Следовательно, применение МНК не позволит добиться требуемой точности

Дисперсии и, значит, стандартные ошибки этих коэффициентов будут смещены (занижены или завышены). Следовательно, применение t- и F-статистик при проверке статистической значимости результатов оценивания, а также построение интервальных оценок параметров становится невозможным.

Рисунок 6.2 - Основные последствия гетероскедастичности

Таким образом, понятно, что гетероскедастичность необходимо выявлять и принимать меры для ее устранения.

## 6.2 Тестирование гетероскедастичности

Обнаружение гетероскедастичности в каждом конкретном случае является довольно сложной задачей, так как для знания дисперсий отклонений  $\sigma^2(\varepsilon_i)$  необходимо знать распределение случайной величины  $Y$ , соответствующее выбранному значению  $x_i$  случайной величины  $X$ . На практике часто для каждого конкретного значения  $x_i$  определяется единственное значение  $y_i$ , что не позволяет оценить дисперсию  $Y$  для данного  $x_i$ . Однако к настоящему времени для такой проверки разработано довольно большое число тестов и критериев

для них. Рассмотрим наиболее популярные и наглядные: графический анализ отклонений, тест Голдфелда– Квандта, тест ранговой корреляции Спирмена, тест Глейзера, тест Парка, тест Уайта и др.

### *Графический анализ остатков*

Использование графического представления отклонений позволяет определиться с возможным наличием гетероскедастичности. В этом случае по оси абсцисс откладываются значения  $x_i$  объясняющей переменной  $X$  (либо линейной комбинации объясняющих переменных  $Y = b_0 + b_1x_1 + \dots + b_px_p$ ), а по оси ординат либо отклонения  $e_i$  либо их квадраты  $e_i^2$ ,  $i = 1, 2, \dots, n$ . Примеры таких графиков приведены на рис. 6.3.

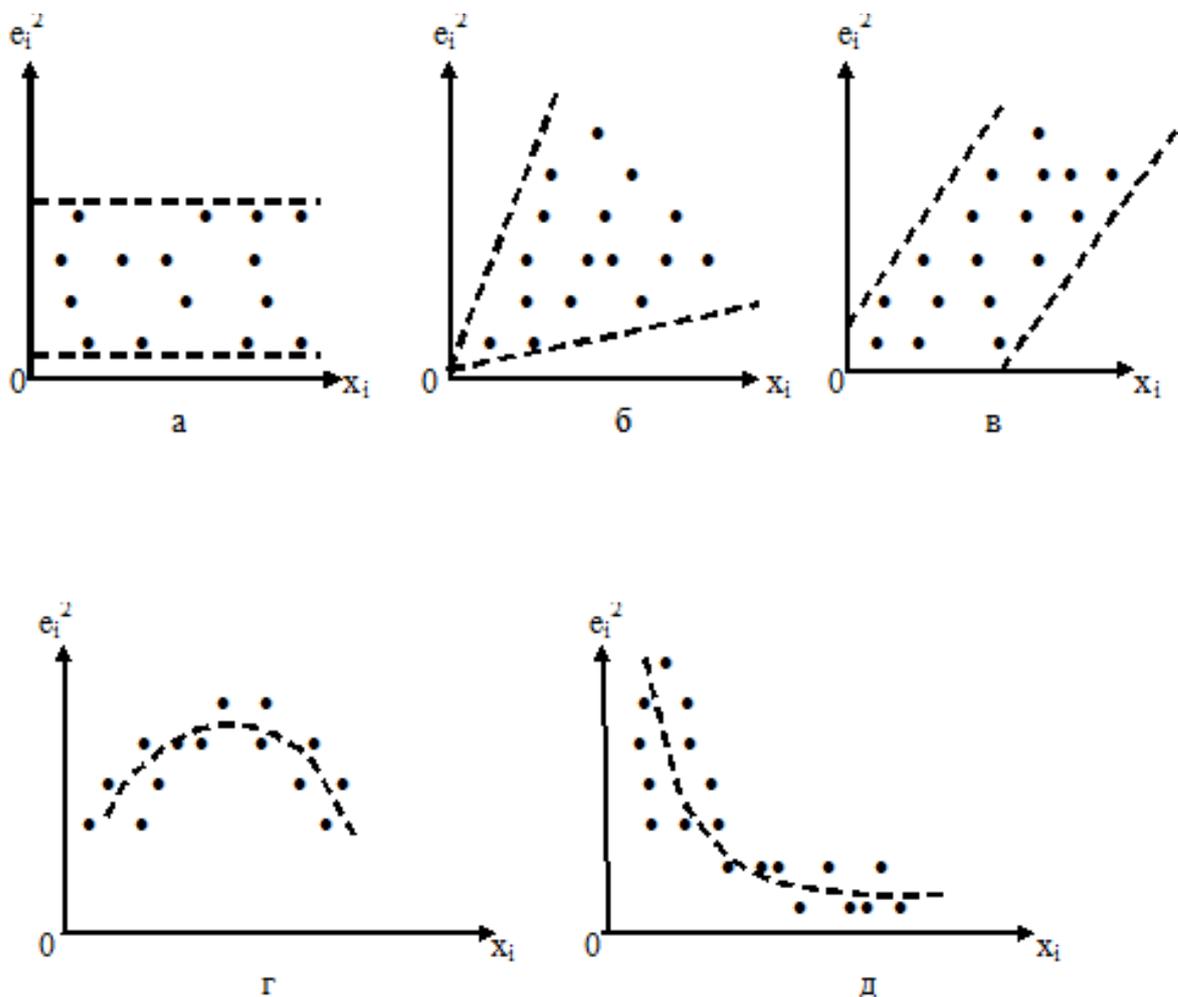


Рисунок 6.3 - Иллюстрация гомоскедастичности и гетероскедастичности

На рис. 6.3а все отклонения  $e_i^2$  находятся внутри полуполосы постоянной ширины, параллельной оси абсцисс. Это говорит о независимости дисперсий  $e_i^2$  от значений переменной  $X$ , т.е. о ее постоянстве, и в этом случае выполняются условия гомоскедастичности.

На рис. 6.3, б – наблюдаются некоторые систематические изменения в соотношениях между значениями  $x_i$  переменной  $X$  и квадратами отклонений  $e_i^2$ .

Рис. 6.3,б соответствует примеру из рисунка 6.1. На рис. 6.3,в отражена линейная, 6.3,г – квадратичная, 6.3,д – гиперболическая зависимости (между квадратами отклонений и значениями объясняющей переменной  $X$ ).

Таким образом, ситуации, представленные на рис. 6.3,б – д, отражают большую вероятность наличия гетероскедастичности для рассматриваемых статистических данных.

Однако, во многих случаях гетероскедастичность не столь очевидна визуально.

Графический анализ остатков, безусловно, является удобным в случае парной регрессии. В случае множественной регрессии подобный анализ необходимо проводить для каждой объясняющей переменной по отдельности, при большом количестве объясняющих переменных вместо объясняющих переменных по оси абсцисс можно отложить значения  $\hat{y}_i$  и графически анализировать наличие гетероскедастичности по графику зависимости  $e_i$  от  $\hat{y}_i$ .

Для того чтобы выявить наличие гетероскедастичности, обычно используют тесты, в которых в качестве нулевой гипотезы рассматривают *гипотезу об отсутствии гетероскедастичности (гипотезу о гомоскедастичности)*.

### Тест Голдфельда-Квандта.

*При проведении проверки по этому критерию предполагается, что стандартное отклонение ( $\sigma_i$ ) распределения вероятностей  $\varepsilon_i$  пропорционально значению  $X$  в этом наблюдении.*

*Предполагается также, что возмущение распределено нормально и не подвержено автокорреляции*

В соответствии с тестом Голдфельда-Квандта выполняем следующие действия (рис. 6.4):

1) все  $n$  наблюдений упорядочим по возрастанию  $X$ ;

2) отбросим  $c$  средних наблюдений (для более надежного разделения групп с малыми и большими значениями дисперсий остатков); величину  $c$  можно определять так:  $c \approx \frac{n \cdot 4}{15}$ , причем удобно принимать такое значение  $c$ , которое позволяет иметь два подмассива **одинаковой длины** (после отбрасывания средних наблюдений);

3) оцениваем отдельные регрессии для первых  $m$  и для последних  $m$  наблюдений и находим суммы квадратов остатков в двух указанных регрессиях, которые обозначим  $S_1$  и  $S_2$ , соответственно

4) рассчитаем отношение  $F = S_2 / S_1$ , которое имеет  $F$ -распределение с  $v_1 = v_2 = m - (p + 1)$  степенями свободы, где  $p$  – число объясняющих переменных в регрессионном уравнении

5) сравним расчетное  $F$  с табличным, делаем вывод: о наличии гетероскедастичности, если  $F$  расчетное больше  $F$  табличного; о гомоскедастичности остатков – в противном случае

Рисунок 6.4 - Алгоритм применения теста Голдфельда-Квандта

Если в модели имеется более одной объясняющей переменной, то наблюдения следует упорядочивать по той из них, которая, как предполагается, связана с  $\sigma_i$ .

Метод Голдфельда-Квандта можно применять и при предположении, что  $\sigma_i$  обратно пропорционально значению  $X$ . При этом используется та же процедура, но тестовой статистикой теперь является отношение  $F = S_1 / S_2$ .

## Тест ранговой корреляции Спирмена

**При использовании данного теста предполагается, что дисперсия регрессионных остатков будет либо увеличиваться, либо уменьшаться с увеличением значений  $X$ . Поэтому для регрессии, построенной с помощью МНК, абсолютные величины остатков  $|e_i|$  и значения  $x_i$  случайной величины  $X$  будут коррелированы**

**Ранг** – номер измерения в вариационном ряду. **Вариационный ряд** представляет собой последовательность, полученную в результате расположения в порядке неубывания исходной последовательности значений некоторой переменной. Иначе говоря, **ранг** - номер объекта в упорядоченном множестве аналогичных объектов. Алгоритм применения теста Спирмена показан на рис. 6.5.

1. Значения  $x_i$  и  $|e_i|$  **ранжируются** (упорядочиваются по величинам).

2. Определяется **коэффициент ранговой корреляции**:

$$r_{x,e} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2 - 1)}, \quad (6.1)$$

где  $d_i$  – разность между рангами  $x_i$  и  $e_i$ ,  $i = 1, 2, \dots, n$ ;  $n$  – число наблюдений.

Например, если  $x_{20}$  является 25-м по величине среди всех наблюдений  $X$ , а  $e_{20}$  является 32-м, то  $d_i = 25 - 32 = -7$ .

3. Для проверки нулевой гипотезы об **отсутствии гетероскедастичности, то есть в данном случае – о равенстве нулю истинного значения коэффициента ранговой корреляции Спирмена  $r$**  – определяется значение t-статистики:

$$t = \frac{r_{x,e} \sqrt{n-2}}{\sqrt{1-r_{x,e}^2}} \quad (6.2)$$

которое сравнивается с табличным значением t-статистики Стьюдента при уровне значимости  $\alpha$  и числе степеней свободы  $\nu = n - 2$ .

4. Если наблюдаемое значение t-статистики, вычисленное по формуле (6.2), превышает  $t_{кр} = t_{\frac{\alpha}{2}, n-2}$ , то необходимо отклонить гипотезу о равенстве нулю коэффициента корреляции  $r_{x,e}$ , а следовательно, и об отсутствии гетероскедастичности. В противном случае гипотеза об отсутствии гетероскедастичности не отклоняется.

Рисунок 6.5 - Алгоритм применения теста Спирмена

Если в модели регрессии больше, чем одна объясняющая переменная, то проверка гипотезы может осуществляться с помощью t-статистики для каждой из них отдельно.

**Р. Парк** предложил тест определения гетероскедастичности, дополняющий графический метод некоторыми формальными зависимостями. Предполагается, что дисперсия  $\sigma_i^2 = \sigma^2(e_i)$  является функцией  $i$ -го значения  $x_i$  переменной  $X$ . То есть, в критерии Парка предполагается определенная форма зависимости дисперсии возмущений от значений объясняющей переменной:  $\sigma_i^2 = \sigma^2 x_i^\beta e^{u_i}$ .

Прологарифмировав данное выражение, получим  $\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln x_i + u_i$

Так как дисперсии  $\sigma_i^2$  обычно неизвестны, то их заменяют оценками квадратов отклонений  $e_i^2$ .

Использование теста Парка включает следующие шаги (рис. 6.6):

1. Строится уравнение регрессии:  $\hat{y}_i = b_0 + b_1 x_i$

2. Для каждого наблюдения определяются остатки:  $e_i = y_i - \hat{y}_i$

3. Определяются логарифмы квадратов остатков:  $\ln e_i^2 = \ln(y_i - \hat{y}_i)^2$

4. Строится уравнение регрессии :  $\ln e_i^2 = \alpha + \beta \ln x_i + u_i$ , где  $\alpha = \ln \sigma^2$ .

В случае множественной регрессии зависимость строится для каждой независимой переменной.

5. Проверяется статистическая значимость коэффициента  $\beta$  на основе t-статистики Стьюдента:  $t = \beta/s_\beta$ . Если коэффициент  $\beta$  статистически значим, то это означает наличие связи между  $\ln e_i^2$  и  $\ln x_i$ , т. е. гетероскедастичности. В противном случае гетероскедастичность данного вида считается незначимой (в этом случае следует использовать также и другие тесты для исключения возможной гетероскедастичности иного вида)

Рисунок 6.6 - Алгоритм применения теста Парка

Таким образом, использование в критерии Парка конкретной функциональной зависимости может привести к необоснованным выводам (например, коэффициент  $\beta$  статистически не значим, а гетероскедастичность имеет место). Поэтому критерий Парка дополняется другими тестами.

### **Тест Бреуша – Пагана**

Этот тест применяется в тех случаях, когда априорно предполагается, что дисперсии  $\sigma_i^2$  зависят от нескольких независимых переменных:

$$\sigma_i^2 = k_0 + \sum_{j=1}^p k_j x_{ij}, \quad i = \overline{1, n},$$

где  $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})'$  - вектор (наблюдаемых) независимых переменных,  $k_0, k = (k_1, \dots, k_p)'$  - неизвестные параметры,  $n$  – количество наблюдений.

В соответствии с тестом Бреуша – Пагана следует действовать так (рис. 6.7):

1) оценить регрессию  $Y$  на  $X$  с помощью МНК, найти остатки  $e_i = Y_i - \hat{Y}_i$  и построить оценку  $\hat{\sigma}^2 = (1/n) \sum e_i^2$ ;

2) провести регрессию  $\frac{e_i^2}{\hat{\sigma}^2} = k_0 + \sum_{j=1}^p k_j x_{ij} + \varepsilon_i, \quad i = \overline{1, n}$  и найти для нее объясненную часть дисперсии  $Y$  (т.е. сумму квадратов отклонений расчетных значений от среднего значения, обозначаемую обычно  $RSS$ );

3) вычислить статистику  $\chi^2 = RSS/2$  и сравнить ее с критическим значением  $\chi^2(p)$  при выбранном уровне значимости и  $p$  степенях свободы. Если вычисленная статистика  $\chi^2$  превосходит критическое значение, то гетероскедастичность присутствует.

Рисунок 6.7 - Алгоритм применения теста Бреуша – Пагана

**Тест Уайта** аналогичен тесту Бреуша-Пагана, но имеет более общий характер. В качестве объясняющих во вспомогательной регрессии используются

все объясняющие переменные исходной модели, их квадраты и всевозможные попарные произведения (рис. 6.8).

- 1) оценить регрессию  $Y$  на  $X$  с помощью обычного МНК и найти остатки  $e_i = Y_i - \hat{Y}_i$  и квадраты остатков  $e_i^2 = (Y_i - \hat{Y}_i)^2$ ;

- 2) построить регрессию квадратов остатков на все регрессоры  $X$ , их квадраты и попарные произведения с помощью обычного МНК (вспомогательная модель):

$$e_i^2 = \alpha_0 + \sum_{j=1}^p \alpha_j x_{ij} + \sum_{j=1}^p \beta_j x_{ij}^2 + \sum_{j=1}^p \sum_{k=1}^p \gamma_{jk} x_{ij} x_{ik} + u_i, i = 1, \dots, n,$$

- и найти для нее коэффициент детерминации  $R^2$ ;

- 3) вычислить статистику  $\chi^2 = nR^2$  и сравнить ее с критическим значением  $\chi^2(p)$  при выбранном уровне значимости и  $p$  степенях свободы. Если вычисленная статистика  $\chi^2$  превосходит критическое значение, то гипотеза о гомоскедастичности отклоняется, и остатки можно считать гетероскедастичными.

Рисунок 6.8 - Алгоритм применения теста Уайта

Привлекательной чертой теста Уайта является его универсальность (он может применяться в любых ситуациях). Однако если гипотеза  $H_0$  отвергается, то этот тест не дает указания на функциональную форму гетероскедастичности.

Кроме того, даже при небольшом количестве независимых переменных в исходной модели во вспомогательной модели может быть много (относительно количества наблюдений) оцениваемых параметров, что существенно снижает мощность теста.

### **Тест Глейзера**

Как некоторые дополнительные предположения относительно характера гетероскедастичности используют **тест Глейзера**. Этот тест дает информацию о форме гетероскедастичности (и тем самым – о способе преобразования модели с целью преодоления гетероскедастичности).

Тест Глейзера во многом аналогичен тесту Уайта, только в качестве зависимой переменной при изучении остатков рассматривается абсолютная

величина остатков, причем в качестве функциональной зависимости используется зависимость вида  $|e_i| = \alpha_0 + \alpha_1 x_i^\gamma + u_i$

Предполагается, что дисперсия возмущения  $\sigma_i^2 = \sigma^2(\varepsilon_i)$ ,  $i = \overline{1, n}$  является некоторой функцией  $i$ -го значения  $(x_i)$  объясняющей переменной  $x$ :

$$\sigma_i = \sigma(\varepsilon_i) = \alpha_0 + \alpha_1 x_i^\gamma + u_i.$$

Чтобы использовать данный метод, следует:

1) оценить регрессионную зависимость  $Y$  от  $X$  с помощью обычного МНК и найти остатки для каждого наблюдения:

$$e_i = Y_i - \hat{Y}_i;$$

2) оценить регрессионную зависимость модулей остатков  $|e_i|$  от независимой переменной  $x$ . Фактическая форма этой регрессии обычно неизвестна, поэтому можно подбирать разные формы кривых, меняя величину показателя степени  $\gamma$  в уравнении

$$|e_i| = \alpha_0 + \alpha_1 x_i^\gamma + u_i,$$

Следует выбрать наиболее подходящую форму, рассматривая коэффициенты корреляции и стандартные ошибки оценок параметров  $a_0$  и  $a_1$ . (При этом, если  $a_0 = 0$  и  $a_1 \neq 0$ , то такая ситуация называется «чистая гетероскедастичность»; если  $a_0 \neq 0$  и  $a_1 \neq 0$ , то этот случай представляет «смешанную гетероскедастичность»).

Гетероскедастичность выявляется при проверке статистической значимости коэффициентов  $a_0$  и  $a_1$ : если они значимо отличны от нуля, то возмущение гетероскедастично.

Если при оценивании более, чем одной функции получается статистически надежная оценка  $a_1$ , то ориентиром при определении характера гетероскедастичности может служить наилучшая из них. Кроме того, критерием выбора значения  $\gamma$  среди рассмотренных значений может быть максимальный коэффициент детерминации  $R^2$ .

Заметим, что в случае многофакторной модели регрессия  $|e_i|$  может строиться не по одной объясняющей переменной, а по нескольким, или по определенной комбинации всех объясняющих переменных (например, по  $x_i = x_{i1}^2 + x_{i2}^2 + \dots + x_{ip}^2$ ) или по  $\hat{Y}$ , представляющей линейную комбинацию всех независимых переменных.

**Недостатком** рассмотренных методов является то, что факт не выявления ими гетероскедастичности, вообще говоря, не означает ее отсутствия: невозможность отклонения гипотезы гомоскедастичности означает в этом случае лишь отсутствие определенного вида зависимости дисперсии возмущений от значений факторов.

### Вопросы для самоконтроля

1. В чем сущность гетероскедастичности?
2. Что является основной причиной гетероскедастичности?
3. Как отражается наличие гетероскедастичности на результатах оценивания регрессионной модели?
4. Какие тесты позволяют определить гетероскедастичность?
5. Приведите аргументы в пользу различных тестов.
6. Каково значение графического теста?
7. В чем суть теста Спирмена?
8. Каковы особенности теста Парка?
9. Приведите алгоритм тестов Бреуша-Пагана и Уайта.
10. Перечислите этапы проверки наличия гетероскедастичности с помощью теста Голдфельда-Квандта.
11. В чем суть и значение теста Глейзера?

## ТЕМА 7. ЛИНЕЙНЫЕ МОДЕЛИ С АВТОКОРРЕЛИРОВАННЫМИ ОСТАТКАМИ. ТЕСТИРОВАНИЕ АВТОКОРРЕЛЯЦИИ

### 7.1 Понятие автокорреляции

Одна из основных предпосылок классической регрессии, выполнение которой является обязательным условием для того, чтобы с помощью МНК получить качественную модель, предусматривает отсутствие автокорреляции возмущений. Если два или больше последовательных уровней ряда остатков коррелированы, *то говорят, что регрессионные остатки подвержены автокорреляции или сериальной корреляции.*

Автокорреляция обычно имеет место, когда используются данные временного ряда. Поэтому в качестве индексации номеров наблюдений используют обычно  $t$ , тем самым подчеркивая, что теперь порядок наблюдений действительно имеет значение.

Таким образом, *автокорреляция возмущений* – это взаимосвязь случайных отклонений (остатков) в различных наблюдениях. Зависимость между остатками описывается с помощью авторегрессионной схемы. Например, допустим, что остаток  $e_t$  находится под влиянием остатка из предыдущего периода времени  $e_{t-1}$  и какого-либо текущего значения случайной переменной  $u_t$ . Остаток  $e_t$  будет описываться следующей авторегрессионной функцией первого порядка:

$$e_t = \rho e_{t-1} + u_t, \quad (7.1)$$

где  $e_t$  - остатки регрессии, полученные обычным методом наименьших квадратов,  
 $u_t$  – случайные переменные, отвечающие всем предпосылкам МНК;  
 $\rho$  – коэффициент авторегрессии (автокорреляции) первого порядка.

Выборочный коэффициент автокорреляции первого порядка равняется:

$$r(1) = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}. \quad (7.2)$$

Основными причинами автокорреляции возмущений являются ошибки спецификации (не все факторы, оказывающие сильное влияние на исследуемый показатель, введены в модель; или неправильно выбрана форма связи; кроме того, может проявляться влияние нескольких второстепенных факторов, не включенных в уравнение и оказывающих однонаправленное воздействие на зависимую переменную). Введение переменных с лагом тоже может привести к автокорреляции остатков.

Применение МНК для оценивания параметров модели при наличии автокорреляции имеет те же негативные последствия, как и в случае с гетероскедастичностью: оценки параметров модели больше не обладают свойством эффективности (хотя могут оставаться состоятельными и несмещенными); стандартные ошибки этих коэффициентов будут смещены (чаще всего занижены). Поэтому применение *t*- и *F*- статистик при проверке статистической значимости результатов оценивания, а также построение интервальных оценок параметров, становится невозможным, что снижает прогнозные качества модели.

## 7.2 Тестирование автокорреляции

Для выявления автокорреляции возмущений используются разные подходы, основанные на изучении остаточных величин (статистических оценок возмущений). Рассмотрим некоторые из них.

### *Графический метод.*

График остатков показывает, что остатки (которые должны быть случайными, не коррелировать друг с другом и иметь одинаковую дисперсию)

ведут себя не совсем адекватно: есть области со строго положительными остатками и области со строго отрицательными остатками (т. е. остатки зависят друг от друга, например, если в каком-то наблюдении остаток положителен, то в следующем наблюдении он, скорее всего, тоже будет положительным). Это визуальный признак автокорреляции. Графики остатков в случае **положительной автокоррелированности** и в случае **отрицательной автокоррелированности** приведены на рис. 7.1, а и б.

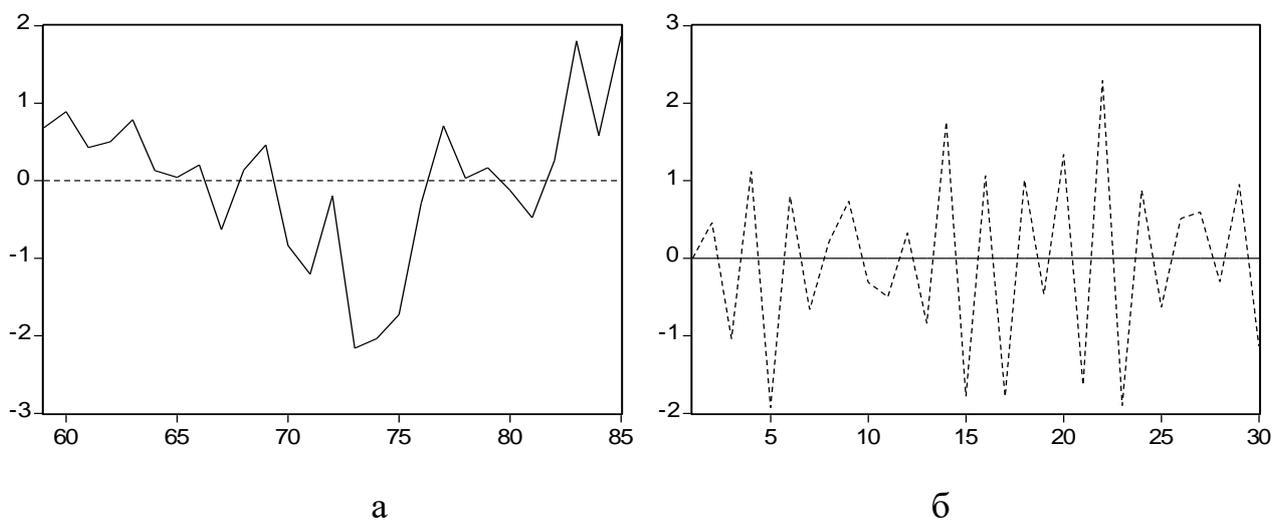


Рисунок 7.1- Положительная (а) и отрицательная(б) автокорреляция остатков

В первом случае (рис. 7.1, а) проявляется **тенденция сохранения знака** остатка при переходе к следующему наблюдению (за положительным остатком следует также положительный остаток, а за отрицательным — отрицательный).

Во втором случае (рис. 7.1, б) проявляется **тенденция смены знака** остатка при переходе к следующему наблюдению (за положительным остатком скорее следует отрицательный остаток, а за отрицательным — положительный).

### **Критерий Дарбина–Уотсона**

Одним из самых популярных тестов проверки на автокорреляцию в эконометрике является тест Дарбина–Уотсона, в соответствии с которым рассчитывается d-статистика (или DW):

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}, \quad (7.3)$$

где  $e_t$  и  $e_{t-1}$  – соответственно текущий и предшествующий уровни ряда остатков, полученных по модели, оцененной с помощью МНК.

Выполним следующие преобразования (7.4):

$$\sum_{t=2}^n (e_t - e_{t-1})^2 = \sum_{t=1}^n e_t^2 - e_1^2 - 2 \sum_{t=2}^n e_t e_{t-1} + \sum_{t=2}^{n+1} e_{t-1}^2 - e_n^2, \quad (7.4)$$

Формулу (7.3) с использованием (7.4) можем записать в виде (7.5):

$$d = \frac{\sum_{t=1}^n e_t^2 + \sum_{t=2}^{n+1} e_{t-1}^2}{\sum_{t=1}^n e_t^2} - \frac{2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} - \frac{e_1^2 + e_n^2}{\sum_{t=1}^n e_t^2} \approx 2 - 2r \quad (7.5)$$

При больших  $n$  последнее слагаемое в (7.5) стремится к нулю.

Таким образом, с коэффициентом автокорреляции  $\rho$  (см. уравнение (7.1))

$d$ -статистика Дарбина–Уотсона связана соотношением:

$$d \approx 2(1 - r),$$

из которого можно получить альтернативную оценку для коэффициента  $\rho$ :

$$r \approx 1 - \frac{d}{2}. \quad (7.6)$$

Очевидно, что если выборочный коэффициент корреляции  $r$  окажется близким к нулю (отсутствии автокорреляции), то значение  $d$  будет близко к двум. Близость  $d$  к нулю свидетельствует о наличии положительной автокорреляции (при этом  $r = 1$ ), к четырем – отрицательной ( $r = -1$ ). Отсюда следует, что  $d$ -статистика может принимать значения в интервале от 0 до 4.

Для статистики Дарбина-Уотсона существуют два пороговых (критических) значения: нижнее  $d_L$  и верхнее  $d_U$  (или  $d_H$  и  $d_B$ ), зависящие от числа наблюдений, числа независимых переменных и уровня значимости критерия (0,01; 0,025; 0,05), значения которых находятся из специальных таблиц (см. табл. А.3 и А.4 Приложения А).

Решение о наличии или отсутствии автокорреляции принимается по следующим правилам. Весь диапазон возможных значений показателя  $d$  разбивается на следующие области принятия решений (рис.7.2):

- 1)  $d_U < d < 4 - d_U$  — область принятия нулевой гипотезы ( $H_0$ ) об отсутствии автокорреляции;
- 2)  $d_L < d < d_U$  или  $4 - d_U < d < 4 - d_L$  — области неопределенности;
- 3)  $0 < d < d_L$  — область принятия альтернативной гипотезы о наличии положительной автокорреляции;
- 4)  $4 - d_L < d < 4$  — область принятия альтернативной гипотезы об отрицательной автокорреляции.

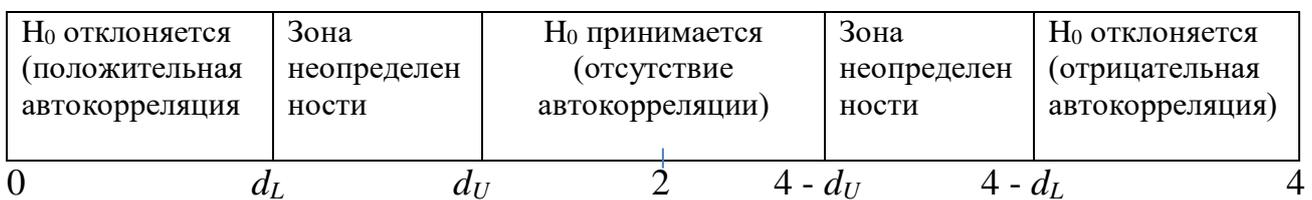


Рисунок 7.2 – Области принятия решений о наличии или отсутствии автокорреляции

Для практического использования удобно применять следующее правило: вычисленное значение  $d$ -статистики сравнивается с табличными, выбранными при заданном уровне значимости, в зависимости от количества наблюдений и числа независимых переменных в модели. Для  $d < 2$  руководствуемся следующим правилом:

- если  $d < d_L$ , то имеется положительная автокорреляция;
- если  $d > d_U$ , то автокорреляции нет;
- если  $d_L < d < d_U$ , то ничего определенного сказать нельзя.

Если расчетное значение  $d$  больше двух, то описанной проверке подвергается величина  $(4 - d)$  и делаются те же выводы с той разницей, что автокорреляция будет отрицательной.

Статистика Дарбина–Уотсона, безусловно, является наиболее важным индикатором наличия автокорреляции. Однако тест обладает и определенными недостатками: 1) наличие зоны неопределенности; 2) ограниченность результата (выявляется лишь корреляция между соседними членами).

Таким образом, есть несколько существенных ограничений на применение критерия Дарбина–Уотсона.

Во-первых, он неприменяется к моделям, включающим в качестве независимых переменных лаговые значения результативного признака, т. е. к моделям авторегрессии.

Во-вторых, методика расчета и использования критерия Дарбина – Уотсона направлена только на выявление автокорреляции остатков первого порядка. При проверке остатков на автокорреляцию более высоких порядков следует применять другие методы.

В-третьих, критерий Дарбина–Уотсона дает достоверные результаты только для больших выборок.

Это приводит к необходимости использовать также и другие тесты. Во всех этих тестах в качестве основной гипотезы  $H_0$  фигурирует гипотеза об отсутствии автокорреляции.

Для устранения недостатков статистики Дарбина-Уотсона Бреуш и Годфри разработали общий тест для обнаружения автокорреляции, который может применяться для высоких порядков авторегрессии случайных отклонений. Тест является асимптотическим, то есть для достоверности выводов требуется большой объем выборки.

**Тест серий (Бреуша– Годфри)** основан на следующем положении [9]: если имеется корреляция между соседними наблюдениями, то естественно ожидать, что в уравнении (7.1) коэффициент  $\rho$  окажется значимо отличающимся от нуля.

Практическое применение теста заключается в оценивании методом наименьших квадратов уравнения регрессии и проверки статистической значимости оценки коэффициента регрессии  $\hat{\rho}$ .

Преимущество теста Бреуша - Годфри по сравнению с тестом Дарбина-Уотсона заключается, во-первых, в том, что он не содержит зоны неопределенности, во-вторых, дает возможность обобщения: в число регрессоров могут быть включены не только остатки с лагом 1, но и с лагом 2, 3 и т.д., что позволяет выявить корреляцию не только между соседними, но и между более отдаленными наблюдениями. Поэтому его называют также тестом серий.

Например, авторегрессионная зависимость остатков от их предыдущих значений представлена следующим уравнением:

$$e_t = 0,56e_{t-1} - 0,12e_{t-2} - 0,01e_{t-3}. \quad (56) \quad (1) \quad (0,10)$$

Величина t-статистик показывает, что значимым оказывается только коэффициент при  $e_{t-1}$ , т.е. существенное влияние на результат наблюдения  $e_t$  оказывает только одно предыдущее значение  $e_{t-1}$ . Положительность оценки соответствующего коэффициента регрессии указывает на положительную корреляцию между ошибками регрессии  $e_t$  и  $e_{t-1}$ .

В более общем виде критерий Бреуша – Годфри представляется следующим образом.

В отличие от автокорреляции первого порядка, оцениваемой критерием Дарбина – Уотсона, рассматривается автокорреляция остатков  $k$ -го порядка:

$$e_t = a_1e_{t-1} + a_2e_{t-2} + \dots + a_ke_{t-k} + u_t \quad (7.7)$$

и строится вспомогательная регрессия МНК-остатков на все независимые переменные, а также случайные отклонения периодов  $t-1, t-2, \dots, t-k$ :

$$e_t = b_1x_1 + b_2x_2 + \dots + b_px_p + a_1e_{t-1} + a_2e_{t-2} + \dots + a_ke_{t-k} + v_t. \quad (7.8)$$

Напомним, что ряд  $e_{t-k}$  представляет ряд  $e_t$  со сдвигом во времени на  $k$  единиц.

В тесте Бреуша – Годфри проверяется нулевая гипотеза:

$$H_0: a_1 = a_2 = \dots = a_k = 0,$$

которая подразумевает отсутствие серийной корреляции какого-либо порядка.

Статистика критерия Бреуша – Годфри равна  $n \cdot R^2$ , где  $R^2$  – коэффициент детерминации, получаемый при оценивании вспомогательной модели (7.8);  $n$  – количество наблюдений, которое должно быть большим.

Нулевая гипотеза отвергается, если вычисленное значение  $n \cdot R^2$  превышает критическое (табличное) значение  $\chi^2$  при заданном уровне значимости  $\alpha$  и  $k$  степенях свободы и не отвергается в противном случае.

Алгоритм применения теста Бреуша – Годфри приведен на рис. 7.3:

1) оценить регрессию  $Y$  на  $X$  с помощью обычного МНК и найти остатки

$$e_t = Y_t - \hat{Y}_t ;$$

2) построить регрессию квадратов остатков на все независимые переменные, а также случайные отклонения периодов  $t-1, t-2, \dots, t-k$ :

$$e_t = b_1 x_1 + b_2 x_2 + \dots + b_p x_p + a_1 e_{t-1} + a_2 e_{t-2} + \dots + a_k e_{t-k} + v_t$$

и найти для нее коэффициент детерминации  $R^2$ ;

3) вычислить статистику  $\chi^2 = nR^2$  и сравнить ее с критическим значением  $\chi^2(k)$  при выбранном уровне значимости  $\alpha$  и  $k$  степенях свободы. Если вычисленная статистика  $\chi^2$  превосходит критическое значение, то гипотеза об отсутствии автокорреляции отклоняется.

Рисунок 7.3 - Алгоритм теста Бреуша – Годфри

### Вопросы для самоконтроля

1. Что такое автокорреляция?
2. Назовите основные причины автокорреляции остатков.
3. В чем отличие положительной и отрицательной автокорреляции?
4. Какая предпосылка МНК нарушается при автокорреляции?
5. Каковы последствия автокорреляции остатков с точки зрения качества модели?
6. Перечислите основные методы обнаружения автокорреляции.
7. Как можно обнаружить автокорреляцию графическим методом?
8. Опишите схему использования статистики Дарбина – Уотсона.
9. Опишите алгоритм применения теста Бреуша-Годфри.
10. Как можно найти эффективные оценки регрессионных коэффициентов в случае линейной модели с коррелированными остатками?
11. Какой критерий используется для проверки гипотезы об автокоррелированности соседних регрессионных остатков?

**ТЕМА 8. ОБОБЩЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИОННАЯ  
МОДЕЛЬ. МЕТОД ЭЙТКЕНА**

### 8.1 Понятие обобщенной линейной регрессионной модели (ОЛРМ)

При моделировании экономических процессов некоторые предпосылки относительно возмущений, сделанные в классической регрессионной модели, являются нереальными. Чаще всего нарушаются условия взаимной некоррелированности и гомоскедастичности возмущений, которые представляются в виде:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Sigma_{\boldsymbol{\varepsilon}} = Cov(\varepsilon_i, \varepsilon_j) = \sigma^2 \mathbf{I}. \quad (8.1)$$

Эти условия оправданы в тех случаях, когда наблюдаемые объекты однородны и описывающие их величины не зависят от своих собственных значений, относящихся к прошлому времени.

Модели регрессии, для которых не выполнены указанные условия, называются **обобщенными линейными регрессионными моделями (ОЛРМ)** (в отличие от классической линейной регрессионной модели (КЛРМ)).

Ковариационная матрица возмущений линейной регрессионной модели имеет вид:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \left[ Cov(\varepsilon_i, \varepsilon_j) \right] = \begin{bmatrix} \sigma_1^2 & Cov(\varepsilon_1, \varepsilon_2) & \cdots & Cov(\varepsilon_1, \varepsilon_n) \\ Cov(\varepsilon_2, \varepsilon_1) & \sigma_2^2 & \cdots & Cov(\varepsilon_2, \varepsilon_n) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(\varepsilon_n, \varepsilon_1) & Cov(\varepsilon_n, \varepsilon_2) & \cdots & \sigma_n^2 \end{bmatrix} = \Omega$$

Эта матрица на главной диагонали содержит дисперсии возмущений различных наблюдений, а внедиагональные элементы представляют их ковариации.

Пусть  $\Omega$  – некоторая симметричная положительно определенная матрица порядка  $n \times n$  (симметричная матрица – квадратная матрица, у которой  $a_{ij} = a_{ji} \forall i, j$ ), где  $n$  – число исходных статистических данных. И пусть ковариационная матрица возмущений ( $\Sigma_\varepsilon$ ) выражается через  $\Omega$  соотношением:

$$\Sigma_\varepsilon = \sigma^2 \Omega .$$

Предполагается, что число  $\sigma_\varepsilon^2$  неизвестно, а матрица  $\Omega$  известна.

**Обобщенная линейная регрессионная модель (ОМЛР)** описывается следующей системой соотношений:

$$\left. \begin{array}{l} 1) \mathbf{Y} = \mathbf{X}\beta + \varepsilon, \\ 2) E(\varepsilon) = \mathbf{0}, \\ 3) E(\varepsilon\varepsilon') = \Sigma_\varepsilon = \sigma^2 \Omega, \\ 4) x_1, x_2, \dots, x_p \text{ – переменные, не случайные переменные,} \\ 5) \text{rang}(\mathbf{X}) = p + 1 < n, \end{array} \right\} \quad (8.2)$$

где  $n$  – количество наблюдений,

$p$  – количество независимых переменных,

$\sigma_\varepsilon^2$  – дисперсия возмущений (неизвестный параметр),

$\Omega$  – известная матрица, которая может быть недиагональной и с различными по величине элементами на главной диагонали.

Сравнение (8.2) с классической регрессионной моделью показывает, что ОМЛР отличается от КЛРМ только видом ковариационной матрицы возмущений, которая представлена формулой (8.1). В КЛРМ матрица  $\Sigma_\varepsilon$  с точностью до неизвестной положительной константы  $\sigma^2$  равна единичной матрице  $\mathbf{I}$ , тогда как в ОМЛР на месте единичной матрицы в описании ковариационной матрицы возмущений появилась матрица  $\Omega$ , то есть теперь дисперсии и ковариации остатков могут быть произвольными при условии невырожденности матрицы  $\Omega$ ,

а в КЛРМ остатки должны быть некоррелированными и гомоскедастичными. Именно в этом и состоит *обобщение* модели.

Оценивание параметров  $\beta_j$  обобщенной линейной регрессионной модели методом наименьших квадратов приведет к следующему:

- 1) оценки  $b_j$  не будут обладать свойством эффективности (хотя останутся несмещенными и состоятельными);
- 2)  $s_e^2$  – оценка дисперсии возмущений является смещенной;
- 3) ковариационная матрица оценок  $\text{Var}(\mathbf{B}) = s_e^2(\mathbf{X}'\mathbf{X})^{-1}$  является смещенным оценщиком истинной ковариационной матрицы. Это означает, что  $t$ -статистики окажутся искаженными, и будет получено неправильное представление о статистической надежности результатов оценивания уравнения регрессии.

Таким образом, обобщенная модель линейной регрессии отличается от классической отказом от требования некоррелированности и гомоскедастичности возмущения и не может быть корректно оценена методом наименьших квадратов.

Для обобщенной регрессионной модели, в отличие от классической, коэффициент детерминации не может служить удовлетворительной мерой качества подгонки. В общем случае он даже не обязан лежать в интервале  $[0, 1]$ , а добавление или удаление независимой переменной не обязательно приводит к его увеличению или уменьшению.

Что можно сделать в случае невыполнения предпосылок?

1. Попытаться изменить спецификацию модели, чтобы устранить автокорреляцию или гетероскедастичность возмущений (например, ввести еще объясняющие переменные).
2. Выбрать такой метод оценивания параметров модели, который при нарушении рассматриваемых предпосылок мог бы максимально обеспечить требуемые свойства полученных оценок, например, обобщенный метод наименьших квадратов.

## 8.2 Обобщенный метод наименьших квадратов (ОМНК)

Сущность *обобщенного метода наименьших квадратов (или метода Эйткена)* состоит в том, чтобы устранить нарушения предпосылок МНК, "скорректировав" расчеты параметров уравнения регрессии с учетом значений ковариационной матрицы остатков. Такая "корректировка" может быть проведена с использованием формулы (8.3).

$$\mathbf{B}_{\text{ОМНК}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{Y}. \quad (8.3)$$

Оценки, найденные с помощью оценщика (8.3), являются несмещенными линейными оценками параметров  $\beta$ , обладающими наименьшей дисперсией (т.е. обладающие свойством эффективности) (доказательству этого посвящена теорема Эйткена).

Обобщенную модель можно преобразовать в классическую, воспользовавшись тем, что матрица  $\mathbf{\Omega}$  известна. После преобразования возмущения модели  $\varepsilon_{\text{пр}}$  должны быть гомоскедастичными и некоррелированными, т.е.

$$\Sigma_{\varepsilon_{\text{пр}}} = \sigma^2\mathbf{I}.$$

Согласно известному результату из матричной алгебры существует такая невырожденная матрица  $\mathbf{P}_{(n \times n)}$ , что:

$$\mathbf{\Omega} = \mathbf{P}\mathbf{P}' \quad (8.4)$$

Умножив (8.4) слева на  $\mathbf{P}^{-1}$ , а справа – на  $(\mathbf{P}')^{-1} = (\mathbf{P}^{-1})'$ , получим:

$$\mathbf{P}^{-1}\mathbf{\Omega}(\mathbf{P}^{-1})' = \mathbf{P}^{-1}\mathbf{P}\mathbf{P}'(\mathbf{P}')^{-1} = \mathbf{I},$$

То есть

$$\mathbf{P}^{-1}\mathbf{\Omega}(\mathbf{P}^{-1})' = \mathbf{I}. \quad (8.5)$$

Если же к (8.4) применить правило обращения произведения квадратных матриц:  $(AB)^{-1} = B^{-1}A^{-1}$ , то из  $\Omega = PP'$  получим:

$$\begin{aligned}\Omega^{-1} &= (PP')^{-1} \\ \Omega^{-1} &= (P')^{-1}P^{-1} \\ \Omega^{-1} &= (P')^{-1}P^{-1} \\ \Omega^{-1} &= (P^{-1})'P^{-1}\end{aligned}\quad (8.6)$$

Теперь обе части уравнения ОЛРМ  $Y = X\beta + \varepsilon$  умножим слева на матрицу  $P^{-1}$ :

$$P^{-1}Y = P^{-1}X\beta + P^{-1}\varepsilon,$$

или 
$$Y_{\text{пр}} = X_{\text{пр}}\beta + \varepsilon_{\text{пр}}, \quad (8.7)$$

где 
$$Y_{\text{пр}} = P^{-1}Y, \quad X_{\text{пр}} = P^{-1}X, \quad \varepsilon_{\text{пр}} = P^{-1}\varepsilon. \quad (8.8)$$

Нетрудно проверить, что модель (8.7) удовлетворяет всем требованиям классической регрессии. Для этого достаточно убедиться в том, что

$$\Sigma_{\varepsilon_{\text{пр}}} = E(\varepsilon_{\text{пр}}\varepsilon_{\text{пр}}') = \sigma^2\mathbf{I}.$$

Действительно,

$$\begin{aligned}\Sigma_{\varepsilon_{\text{пр}}} &= E(\varepsilon_{\text{пр}}\varepsilon_{\text{пр}}') = E[(P^{-1}\varepsilon)(P^{-1}\varepsilon)'] = E[P^{-1}\varepsilon\varepsilon'(P^{-1})'] = \\ &= P^{-1}E[\varepsilon\varepsilon'](P^{-1})' = P^{-1}\sigma^2\Omega(P^{-1})' = \sigma^2\underbrace{[P^{-1}\Omega(P^{-1})']}_{\text{Из (6.5)}} = \sigma^2\mathbf{I}\end{aligned}$$

Применив к преобразованным данным МНК, получим формулу (8.9) – оператор оценивания для обобщенного метода наименьших квадратов:

$$\begin{aligned}\mathbf{B} &= [(P^{-1}\mathbf{X})'(P^{-1}\mathbf{X})]^{-1}(P^{-1}\mathbf{X})'(P^{-1}\mathbf{Y}) = [\mathbf{X}'\underbrace{(P^{-1})'P^{-1}}_{\Omega^{-1}}\mathbf{X}]^{-1}\mathbf{X}'\underbrace{(P^{-1})'P^{-1}}_{\Omega^{-1}}\mathbf{Y} = \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}\end{aligned}\quad (8.9)$$

Итак, эффективной оценкой (наилучшей в классе всех линейных (относительно  $Y$ ) несмещенных оценок) параметра  $\beta$  будет оценка

$$\mathbf{B} = (\mathbf{X}'_{\text{пр}} \mathbf{X}_{\text{пр}})^{-1} \mathbf{X}'_{\text{пр}} \mathbf{Y}_{\text{пр}},$$

или

$$\mathbf{B}_{\text{ОМНК}} = (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{Y},$$

ковариационная матрица которой определяется формулой:

$$\boldsymbol{\Sigma}_B = \sigma^2 (\mathbf{X}'_{\text{пр}} \mathbf{X}_{\text{пр}})^{-1},$$

или

$$\boldsymbol{\Sigma}_{\mathbf{B}_{\text{ОМНК}}} = \sigma^2 (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}.$$

Такими образом, если преобразовать переменные  $X$  и  $Y$  по формулам (8.8), то применение к ним обычного МНК приведет к тому же результату, что и применение ОМНК к исходным переменным. В этом и состоит суть метода Эйткена со вспомогательной моделью:

- 1) *преобразование обобщенной модели во вспомогательную, удовлетворяющую предпосылкам МНК, по схеме:*

$$\mathbf{Y}_{\text{пр}} = \mathbf{P}^{-1} \mathbf{Y}, \quad \mathbf{X}_{\text{пр}} = \mathbf{P}^{-1} \mathbf{X};$$

- 2) *применение ко вспомогательной модели метода наименьших квадратов. Полученные оценки будут равнозначны оценкам Эйткена для параметров обобщенной модели регрессии.*

Итак, имеются два способа применения ОМНК: 1) с использованием ковариационной матрицы остатков  $\boldsymbol{\Omega}$  и формулы (8.4) и 2) с использованием матрицы  $\mathbf{P}$  и модели с преобразованными переменными (8.6), для решения которой достаточно применить обычный МНК.

Основная проблема, возникающая при использовании ОМНК, заключается в том, что фактические значения элементов матрицы  $\boldsymbol{\Omega}$  неизвестны. Поэтому для применения этого метода используют их оценки, полученные на основе исследования имеющихся в распоряжении данных. В этом случае говорят о *доступном ОМНК*, в соответствии с которым для оценки элементов

ковариационной матрицы остатков  $\Omega$  выдвигают разные предположения об их характере и структуре.

В качестве оценок дисперсий возмущений (диагональных элементов матрицы  $\Omega$ ) могут быть использованы квадраты остатков, полученных при применении обычного МНК, поскольку выполняется предпосылка о равенстве нулю математического ожидания возмущений.

Как правило, считают, что гетероскедастичность остатков присуща данным пространственных выборок, а автокорреляция остатков – временным рядам. Поэтому в зависимости от характера исходных данных часто предполагают или гетероскедастичность остатков при условии отсутствия автокорреляции, или их автокорреляцию при условии гомоскедастичности.

Определим вид матрицы  $P^{-1}$ , которая используется для преобразования исходных данных.

В зависимости от выполнения рассматриваемых основных предпосылок классической регрессии ковариационная матрица имеет тот или иной вид:

- **при гомоскедастичности (постоянстве дисперсий случайной компоненты) и отсутствии автокорреляции (взаимной некоррелированности возмущений)**

$$\Sigma_{\varepsilon} = \begin{bmatrix} \sigma_1^2 & Cov(\varepsilon_1, \varepsilon_2) & \cdots & Cov(\varepsilon_1, \varepsilon_n) \\ Cov(\varepsilon_2, \varepsilon_1) & \sigma_2^2 & \cdots & Cov(\varepsilon_2, \varepsilon_n) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(\varepsilon_n, \varepsilon_1) & Cov(\varepsilon_n, \varepsilon_2) & \cdots & \sigma_n^2 \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I;$$

- **при гетероскедастичности и отсутствии автокорреляции возмущений** ковариационная матрица возмущений по-прежнему диагональная (т.е. все

внедиагональные элементы равны нулю), но с различными по величине диагональными элементами (дисперсиями):

$$\Sigma_{\varepsilon} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} = \text{diag}(\sigma_i^2), \quad i = \overline{1, n}.$$

При этом, учитывая, что  $\Sigma_{\varepsilon} = \sigma^2 \Omega$ , можем записать:

$$\Omega = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

Ранее приняли  $\Omega = \mathbf{P}\mathbf{P}'$ , поэтому  $\Omega^{-1} = (\mathbf{P}^{-1})'\mathbf{P}^{-1}$ .

$$\Omega^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\lambda_n} \end{bmatrix}.$$

Таким образом, *в условиях невыполнения предпосылки о гомоскедастичности возмущений, но при отсутствии автокорреляции*

$$\mathbf{P}^{-1} = \sqrt{\Omega^{-1}} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\lambda_n}} \end{bmatrix}; \quad (8.10)$$

Соответствующие преобразованные исходные данные имеют при такой матрице преобразования  $P^{-1}$  следующий вид:

$$Y_{np} = P^{-1}Y = \begin{pmatrix} \frac{Y_1}{\sqrt{\lambda_1}} \\ \frac{Y_2}{\sqrt{\lambda_2}} \\ \frac{Y_3}{\sqrt{\lambda_3}} \\ \dots \\ \frac{Y_n}{\sqrt{\lambda_n}} \end{pmatrix}; \quad X_{np} = P^{-1}X = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & \frac{X_{11}}{\sqrt{\lambda_1}} & \dots & \frac{X_{1p}}{\sqrt{\lambda_1}} \\ \frac{1}{\sqrt{\lambda_2}} & \frac{X_{21}}{\sqrt{\lambda_2}} & \dots & \frac{X_{2p}}{\sqrt{\lambda_2}} \\ \dots & \dots & \dots & \dots \\ \frac{1}{\sqrt{\lambda_n}} & \frac{X_{n1}}{\sqrt{\lambda_n}} & \dots & \frac{X_{np}}{\sqrt{\lambda_n}} \end{pmatrix}. \quad (8.11)$$

Здесь  $\sqrt{\lambda_i}$  – неизвестные значения, которые должны быть оценены. В качестве одного из возможных методов их оценивания можно использовать расчетные значения модуля остатков для «лучшего» уравнения, описывающего зависимость остатков от выбранного фактора, полученные при использовании теста Глейзере (см. п.6.2):  $|\hat{\varepsilon}_i| = a_0 + a_1 X_i^? = \sqrt{\lambda_i}$

- **при автокорреляции (первого порядка) и гомоскедастичности возмущений** матрица ковариаций не может быть диагональной, поскольку в ситуациях, когда исходные наблюдения регистрируются во времени, возмущения оказываются взаимозависимыми, коррелированными.

Естественным допущением относительно природы зависимости регрессионных остатков является гипотеза, в соответствии с которой эта зависимость ослабевает по мере их взаимного удаления друг от друга во времени. Одной из удобных форм реализации этого допущения (при сохранении свойства гомоскедастичности остатков) является следующая:

$$r(\varepsilon_i, \varepsilon_j) = \rho^{|i-j|},$$

где  $r(\varepsilon_i, \varepsilon_j)$  – коэффициент корреляции между  $\varepsilon_i$  и  $\varepsilon_j$ , а  $\rho$  – некоторое число, по модулю меньше единицы. Очевидно, что по своему вероятностному смыслу  $\rho$  – это коэффициент корреляции между соседними по времени остатками, так как

$$\rho^{2-1} = \rho^{3-2} = \dots = \rho^{15-14} = \rho^1.$$

При  $|i-j| \rightarrow \infty$  (т.е. при неограниченном удалении остатков друг от друга во времени) величина  $\rho$  стремится к нулю, (связь исчезает).

$$\text{Ковариация остатков равна: } \sigma_{ij} = \sigma_{\varepsilon}^2 \rho^{|i-j|}.$$

$$\text{Действительно, } \text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij} = r(\varepsilon_i, \varepsilon_j) \cdot \sigma_{\varepsilon_i} \cdot \sigma_{\varepsilon_j} = \rho^{|i-j|} \sigma_{\varepsilon}^2,$$

где  $\sigma_{\varepsilon}^2 = D(\varepsilon|X_i)$  – условная дисперсия остатков (не зависящая от величины  $X_i$  в силу гомоскедастичности). Так что ковариационная матрица остатков имеет в данном случае вид:

$$\Sigma_s = \sigma_{\varepsilon}^2 \begin{pmatrix} \rho^{|1-1|} & \rho^{|1-2|} & \rho^{|1-3|} & \dots & \rho^{|1-n|} \\ \rho^{|2-1|} & \rho^{|2-2|} & \rho^{|2-3|} & \dots & \rho^{|2-n|} \\ \rho^{|3-1|} & \rho^{|3-2|} & \rho^{|3-3|} & \dots & \rho^{|3-n|} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{|n-1|} & \rho^{|n-2|} & \rho^{|n-3|} & \dots & \rho^{|n-n|} \end{pmatrix} = \sigma_{\varepsilon}^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}.$$

Таким образом, при автокорреляции (первого порядка)

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}.$$

Тогда

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}.$$

$$\mathbf{P}^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \quad (8.12)$$

Соответствующие преобразованные исходные данные имеют при такой матрице преобразования  $\mathbf{P}^{-1}$  следующий вид:

$$Y_{np} = P^{-1}Y = \begin{pmatrix} \sqrt{1-\rho^2}Y_1 \\ Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ \dots \\ Y_n - \rho Y_{n-1} \end{pmatrix}; \quad (8.13)$$

$$X_{np} = P^{-1}X = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}X_{11} & \dots & \sqrt{1-\rho^2}X_{1p} \\ 1-\rho & X_{21} - \rho X_{11} & \dots & X_{2p} - \rho X_{1p} \\ \dots & \dots & \dots & \dots \\ 1-\rho & X_{n1} - \rho X_{n-1,1} & \dots & X_{np} - \rho X_{n-1,p} \end{pmatrix}, \quad (8.14)$$

где  $\rho$  – неизвестный параметр, который следует оценить. Его оценку производят, исследуя взаимосвязь остатков, полученных после применения к модели регрессии обычного МНК.

Рассмотрим некоторые методы оценивания.

**1. Оценивание  $\rho$  на основе статистики Дарбина-Уотсона**, которая, как показано ранее, тесно связана с коэффициентом корреляции между соседними отклонениями (см. (7.6):

$$r \approx 1 - \frac{d}{2}.$$

Здесь  $r$  представляет выборочную оценку коэффициента  $\rho$ .

Этот метод неплох при большом числе наблюдений; оценка  $r$  параметра  $\rho$  будет достаточно точной.

**2. Метод (процедура) Кохрейна – Оркатта** (носит итеративный характер).

Рассмотрим модель

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = \overline{1, n}. \quad (8.15)$$

Предположим, что возмущения  $\varepsilon_i$  образуют процесс авторегрессии первого порядка:

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i, \quad i = \overline{1, n},$$

где  $0 < \rho < 1$ , а  $u_i$ ,  $i = \overline{1, n}$  – нормально распределенные случайные величины, не коррелирующие с  $\varepsilon_{i-s}$  для  $s > 0$ .

Умножим обе части уравнения (8.15) для  $(i - 1)$ -го наблюдения на  $\rho$

$$\rho y_{i-1} = \beta_0 \rho + \beta_1 \rho x_{i-1} + \rho \varepsilon_{i-1}, \quad i = \overline{1, n}$$

и вычтем обе части полученного выражения из соответствующих частей выражения (8.15) для  $i$ -го наблюдения:

$$y_i - \rho y_{i-1} = \beta_0(1 - \rho) + \beta_1(x_i - \rho x_{i-1}) + (\varepsilon_i - \rho \varepsilon_{i-1})$$

(*авторегрессионное преобразование*). Таким образом, приходим к преобразованной модели

$$y'_i = \alpha + \beta_1 x'_i + \varepsilon'_i, \quad i = \overline{1, n},$$

где

$$y'_i = y_i - \rho y_{i-1}$$

$$x'_i = x_i - \rho x_{i-1}$$

$$\varepsilon'_i = \varepsilon_i - \rho \varepsilon_{i-1}$$

$$\alpha = \beta_0(1 - \rho)$$

Процедура Кохрейна – Оркатта использует для аппроксимации теоретического преобразования оценку для  $\rho$  в виде:

$$r = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=2}^n e_{i-1}^2}$$

где  $e_i$  - остатки, получаемые при оценивании *исходной* модели.

Если в преобразованной модели автокоррелированность не выявляется, то полученные в рамках этой модели оценки параметров можно принять в качестве уточненных оценок исходных параметров. В противном случае процесс преобразования применяют уже к преобразованной модели и еще раз уточняют значения оценок параметров и т.д., пока последовательно уточняемые оценки не перестанут изменяться в пределах заданной точности.

Получив оценку  $a$  для  $\alpha$ , можно оценить параметр  $\beta_0$  исходной модели:

$$b_0 = \frac{a}{1 - \rho}.$$

Процедура Кохрейна — Оркатта реализована в большинстве эконометрических компьютерных программ.

Таким образом, при обнаружении нарушений стандартных предположений следует либо улучшить спецификацию модели, привлекая подходящие дополнительные объясняющие переменные, либо использовать для оценивания коэффициентов и оценивания дисперсий коэффициентов модели специальные методы оценивания, принимающие во внимание обнаруженные нарушения (например, *взвешенный метод наименьших квадратов* или *авторегрессионное преобразование переменных*).

Так, при установлении автокорреляции необходимо в первую очередь проанализировать правильность спецификации модели.

Если после уточнения состава объясняющих переменных, либо изменения формы зависимости автокорреляции по-прежнему имеет место, то, возможно, это связано с внутренними свойствами ряда остатков. В этом случае возможны определенные преобразования, устраняющие автокорреляцию. Среди них выделяется  $AR(1)$ , которая может быть обобщена в  $AR(k)$ ,  $k = 2, 3, \dots$

### Вопросы для самоконтроля

1. Что понимается под обобщенной линейной регрессионной моделью?
2. Как можно описать обобщенную линейную регрессионную модель?
3. Чем отличается обобщенная линейная регрессионная модель от классической линейной регрессионной модели?
4. В чем сущность обобщенного метода наименьших квадратов (ОМНК)?
5. Почему на практике используются различные модификации обобщенного метода наименьших квадратов?
6. Как с помощью ОМНК со вспомогательной матрицей найти оценки параметров модели при гетероскедастичности остатков?
7. Как с помощью ОМНК со вспомогательной матрицей найти оценки параметров модели при автокорреляции остатков?
8. В чем суть оценивания параметров модели на основе статистики Дарбина-Уотсона? В каком случае такой метод применяется?
9. Для чего и как используется процедура Кохрейна—Оркатта?

## ТЕМА 9. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ (СОУ)

### 9.1 Необходимость использования СОУ

До сих пор при изучении регрессионных зависимостей рассматривались односторонние стохастические причинные отношения между экономическими явлениями и процессами, описываемые уравнением:

$$Y = f(x_1, x_2, \dots, x_p, \varepsilon). \quad (9.1)$$

При этом мы исходили из того, что изменение зависимой переменной  $Y$  объяснялось переменными  $x_1, x_2, \dots, x_p$  и что эти объясняющие переменные (в правой части регрессионного уравнения (9.1)) некоррелированы с возмущающей переменной  $\varepsilon$  (это предположение нашло отражение в четвертой предпосылке МНК). И нас интересовали методы оценивания именно одного уравнения.

Однако в экономике редко можно встретить подобные односторонние стохастические причинные отношения. Чаще всего объектом статистического изучения в социально-экономических науках являются сложные системы, для описания и объяснения механизма функционирования которых недостаточно изолированных уравнений регрессии. При использовании отдельных уравнений в большинстве случаев полагается, что факторы (аргументы) можно изменять независимо друг от друга, тогда как это предположение является очень грубым: изменение одной переменной, как правило, повлечет за собой изменения во всей системе взаимосвязанных признаков.

Поэтому в экономических исследованиях важное место заняла проблема описания структуры связей между показателями системой регрессионных

уравнений, каждое из которых отображает одну из зависимостей, закономерностей изменения, свойств изучаемого сложного объекта.

Система уравнений, в которой одни и те же переменные в разных уравнениях могут одновременно выступать, с одной стороны, в роли результирующих, (объясняемых) переменных, а с другой стороны – в роли объясняющих переменных, называется *системой взаимозависимых, или одновременных уравнений*. Примеры таких систем приведены в табл. 9.1.

## 9.2 Переменные в системах одновременных уравнений

При построении эконометрических моделей недостаточно принятого ранее разделения переменных на объясняющие и зависимую, поскольку одна и та же переменная может входить в одно уравнение как зависимая, а в другое – как объясняющая. Будем различать в эконометрической модели следующие переменные.

**1. Эндогенные переменные.** Эндогенными переменными являются экономические величины, которые объясняются эконометрической моделью. Это переменные, значения которых определяются внутри модели, они формируются в результате одновременного взаимодействия факторов, входящих в модель. Эндогенные переменные зависят от экзогенных и возмущающих переменных.

**2. Экзогенные переменные.** Экзогенные переменные являются внешними, наперед заданными экономическими величинами. Их значения в каждый период времени  $t$  определяются вне модели. Они, следовательно, объясняются не моделью, а экономическими факторами и закономерностями, лежащими за границами этой модели.

Экзогенные переменные определяют изменение эндогенных переменных, но сами не находятся под их влиянием. Таким образом, между эндогенными и экзогенными переменными существуют только односторонние стохастические причинные отношения.

Таблица 9.1 – Примеры систем одновременных уравнений

Название модели	Модель	Описание модели
Спрос – предложение	Функция спроса: $q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{1t}$ , $\alpha_1 < 0$ , Функция предложения: $q_t^S = \beta_0 + \beta_1 p_t + \varepsilon_{2t}$ , $\beta_1 < 0$ , Уравнение равновесия: $q_t^D = q_t^S$ .	Одна из простейших СОУ используется при моделировании спроса – предложения в рыночной экономике. В предположении, что спрос $q^D$ и предложение $q^S$ в момент времени $t$ являются линейными функциями от цены $p$ в этот же момент времени.
	С учетом дохода потребителей ( $Y$ )  Функция спроса: $q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 Y_t + \varepsilon_{1t}$ , $\alpha_1 < 0$ , Функция предложения: $q_t^S = \beta_0 + \beta_1 p_t + \varepsilon_{2t}$ , $\beta_1 < 0$ , Уравнение равновесия: $q_t^D = q_t^S$ .	Очевидно, что наличие случайных переменных в данных уравнениях связано в первую очередь с отсутствием ряда важных объясняющих переменных (дохода, цен, сопутствующих товаров, вкусов, ожиданий, цены ресурсов, налогов и т. д.). Изменение одного из этих факторов может отразиться на модели. Например, рост дохода потребителей может сдвинуть кривую спроса вверх. Это приведет к изменению равновесной цены и равновесного количества.
Кейнсианская модель формирования доходов	Функция потребления: $C_t = \beta_0 + \beta_1 Y_t + \varepsilon_t$ , Макроэкономическое тождество: $Y_t = C_t + I_t$ .	Простейшая модель такого типа в предположении, что рассматривается закрытая экономика без государственных расходов. Здесь $Y_t$ , $C_t$ , $I_t$ представляют совокупный выпуск, объемы потребления и инвестиций в момент времени $t$ соответственно.
Модели IS – LM	<u>Модель IS</u> Функция потребления: $C_t = \beta_0 + \beta_1 Y_{(d)t}$ , Функция налогов: $T_t = \alpha_0 + \alpha_1 Y_t$ , Функция инвестиций: $I_t = \gamma_0 + \gamma_1 r_t$ , Располагаемый доход: $Y_{(d)t} = Y_t - T_t$ , Государственные расходы: $G_t = \bar{G}$ , Макроэкономическое тождество: $Y_t = C_t + I_t + G_t$ .	Модель IS – равновесие на рынке товаров, задающая такое соотношение между процентной ставкой и уровнем дохода, при котором рынок товаров находится в равновесии. Линия LM (линия равновесия на рынке денег) задает такое соотношение между процентной ставкой и уровнем дохода, при котором спрос на деньги равен их предложению. Здесь $Y_t$ , $C_t$ , $I_t$ , $G_t$ , $T_t$ , $Y_{(d)t}$ , $r_t$ – значения в момент времени $t$ соответственно национального дохода, потребления, желаемого объема чистых инвестиций, государственных расходов ( $G_t$ в данном случае $G_t = \bar{G} = \text{const}$ ), объема налогов, располагаемого дохода, процентной ставки.
	<u>Модель LM</u> Функция спроса на деньги: $M_t^D = a + bY_t - cr_t$ , Функция предложения денег: $M_t^S = \bar{M}$ , Условие равновесия: $M_t^D = M_t^S$ .	

Решение, какие из переменных, включенных в модель, следует отнести к экзогенным, а какие – к эндогенным, принимается при каждой спецификации модели. Деление переменных на эндогенные и экзогенные относительно. Оно зависит от природы изучаемого явления, а также от цели, с которой эта модель строится.

### ***3. Предопределенные переменные.***

Эндогенные и экзогенные переменные могут быть также лаговыми. *Под лаговой* понимают переменную, значения которой отстают на один или несколько периодов. Поскольку значения лаговых переменных определены ранее (до рассмотрения уравнения), то в период времени  $t$  они также не объясняются эконометрической моделью, и их можно отнести к заранее заданным экзогенным. В связи с этим к классу *предопределенных* переменных относятся:

- обычные экзогенные переменные; они заранее предопределены, так как объясняются не эконометрической моделью, а факторами, лежащими вне этой модели;
- лаговые экзогенные переменные; они заранее предопределены, так как их значения принадлежат предшествующим периодам и объясняются вне модели;
- лаговые эндогенные переменные; их предопределенность следует из предшествующего объяснения.

К предопределенным относятся и фиктивные переменные, тождественно равные единице и введенные в уравнение регрессии для свободного члена.

Такая классификация переменных позволяет указать методы определения эндогенных переменных. Например, в функции спроса и предложения (табл. 9.1), переменные, определяющие величину спроса  $q_t^D$ , предложения  $q_t^S$ , цену  $p$ , являются эндогенными, так как они определяются внутри системы.

В кейнсианской модели формирования доходов (табл. 9.1)  $C$  и  $Y$  являются эндогенными, которые оцениваются внутри модели, а переменная  $I$ , задаваемая (определяемая) вне модели, является экзогенной. Из модели нельзя понять, как получают значения этой переменной. Они используются как заранее заданные.





в одних уравнениях и как независимые в других. В эконометрике эта система уравнений называется также *структурной формой модели*.

В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания.

**Полная эконометрическая модель.** Эконометрическая модель называется полной, если:

- включенные в нее predetermined переменные оказывают существенное влияние на совместно зависимые переменные, а возмущения носят случайный характер;
- количество уравнений в ней совпадает с числом совместно зависимых переменных, так что каждая совместно зависимая переменная может быть объяснена с помощью соответствующего уравнения;
- система уравнений имеет однозначное решение относительно совместно зависимых переменных. Следовательно, матрица  $A$  (матрица оценок структурных параметров при эндогенных переменных) невырожденная, т.е.  $\det A \neq 0$ .

Модель должна быть полной, если необходимо количественно описать экономическое явление или когда она применяется для прогнозирования.

#### 9.4 Структурная и приведенная формы модели

Уравнения, составляющие исходную модель, называются *структурными уравнениями модели*. Структурная форма эконометрической модели создается в процессе формирования самой модели при стремлении отразить причинно-следственный механизм, существующий в реальности. Она описывает односторонние и многосторонние стохастические причинные отношения между экономическими величинами в их непосредственном виде и содержит всю

существенную информацию о зависимостях между экономическими явлениями и процессами.

*Таким образом, структурная форма эконометрической отражает существующую структуру взаимосвязи между объясняемой эндогенной переменной каждого уравнения и соответствующими объясняющими переменными* (в том или приблизительно в том виде, который диктуется соображениями экономической теории).

Каждое структурное уравнение модели содержит одну или несколько эндогенных (совместно зависимых) переменных, из которых только одна является объясняемой для данного уравнения, а остальные выполняют функции объясняющих переменных. Оно включает неизвестные, подлежащие оцениванию структурные параметры, а также возмущение.

Структурная модель (9.7) в матричном виде:

$$\alpha Y + \beta X + \varepsilon = 0, \quad (9.6)$$

где

$$\alpha = \begin{bmatrix} -1 & \alpha_{12} & \cdots & \alpha_{1m} \\ \alpha_{21} & -1 & \cdots & \alpha_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & -1 \end{bmatrix} \beta = \begin{bmatrix} \beta_{10} & \beta_{11} & \cdots & \beta_{1p} \\ \beta_{20} & \beta_{21} & \cdots & \beta_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{m0} & \beta_{m1} & \cdots & \beta_{mp} \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \cdots \\ Y_{mt} \end{bmatrix} \varepsilon = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \cdots \\ \varepsilon_{mt} \end{bmatrix} \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}$$

Наряду со структурными уравнениями (*поведенческие уравнения*) эконометрическая модель может содержать **тождества**, которые необходимы для адекватного отражения реальности и полного охвата переменных одновременными соотношениями. Тождества не содержат возмущающих переменных, и их параметры известны (в общем случае равны единице). Следовательно, они не подлежат оценке, и при проведении процедуры оценивания параметров модели могут быть заранее исключены.

Уравнения, в которых отражена схема определения эндогенных переменных, называется **уравнениями в приведенной (или редуцированной) форме (или приведенными уравнениями)**.

Установим связь между параметрами структурной и приведенной форм.

Структурную модель (см. (9.7)) запишем, используя оценки параметров, в следующем виде:

$$\mathbf{AY} + \mathbf{BX} + \mathbf{e} = \mathbf{0}.$$

Решим соответствующее матричное уравнение относительно  $\mathbf{Y}$ :

$$\begin{aligned} \mathbf{AY} &= \mathbf{BX} + \mathbf{e}, \\ \mathbf{A}^{-1} \mathbf{A} \mathbf{Y} &= \mathbf{A}^{-1} \mathbf{BX} + \mathbf{A}^{-1} \mathbf{e}, \\ \mathbf{Y} &= \mathbf{A}^{-1} \mathbf{BX} + \mathbf{A}^{-1} \mathbf{e}, \\ \text{или } \mathbf{Y} &= \mathbf{CX} + \mathbf{u}, \end{aligned} \tag{9.7}$$

где  $\mathbf{C} = \mathbf{A}^{-1} \mathbf{B}$ ,  $\mathbf{u} = \mathbf{A}^{-1} \mathbf{e}$ .

Уравнение (9.7) представляет **приведенную (редуцированную) форму эконометрической модели**.

В развернутом виде приведенная модель (9.8):

$$\left\{ \begin{array}{l} Y_1 = c_{10}X_0 + c_{11}X_1 + \dots + c_{1p}X_p + u_1 \\ Y_2 = c_{20}X_0 + c_{21}X_1 + \dots + c_{2p}X_p + u_2 \\ Y_3 = c_{30}X_0 + c_{31}X_1 + \dots + c_{3p}X_p + u_3 \\ \dots \dots \dots \\ Y_m = c_{m0}X_0 + c_{m1}X_1 + \dots + c_{mp}X_p + u_m \end{array} \right. \tag{9.8}$$

Из (9.7) и (9.8) хорошо видно, что **отличительной особенностью приведенной) формы** является **отсутствие эндогенных переменных в качестве объясняющих во всех уравнениях системы**; все объясняющие переменные уравнений этой формы являются **предопределенными переменными системы уравнений**.

Модель в приведенной форме используется для прогнозирования. И напротив, структурная форма эконометрической модели для прогноза непригодна, так как в каждом структурном уравнении содержится несколько совместно зависимых переменных, для которых не могут быть указаны значения на прогнозный период, поскольку они еще только подлежат оценке.

Непосредственное использование МНК для оценки параметров структурных уравнений в большинстве случаев приводит к неудовлетворительному результату (оценки получаются смещенными и несостоятельными, а статистические выводы по ним – некорректными) из-за коррелированности одной или нескольких объясняющих переменных с возмущением, т.е. из-за нарушения одной из основных предпосылок классической регрессии. Зато для оценивания параметров приведенной формы обычный МНК может быть использован.

### **9.5 Проблема идентифицируемости. Условия идентифицируемости уравнений системы**

При анализе эконометрической модели, представленной системой уравнений, исследователя, прежде всего, интересует поведение эндогенных переменных  $Y_t$ . Из соответствующей приведенной формы видно, что эндогенные переменные  $Y_t$  являются по своей природе случайными величинами, поведение которых определяется *внутренней структурой модели*, а именно элементами матриц  $A$  и  $B$  и природой случайных остатков  $e_t$ . Возникает вопрос: а возможно ли, следуя в “обратном направлении”, восстановить структурную форму (т.е. элементы матриц  $A$  и  $B$ ), располагая знанием приведенной формы (т.е. знанием числовых значений всех элементов матрицы  $C$  и природы случайных остатков  $e_t$ )? Именно этот вопрос и отражает *сущность проблемы идентифицируемости* эконометрической модели.

Оказывается – не всегда. Недостатком приведенной формы является как раз то обстоятельство, что от количественно оцененной модели в приведенной форме

не во всех случаях можно перейти к модели в структурной форме, тогда как по заданной в численном виде структурной форме может быть всегда определена приведенная форма.

С позиции идентифицируемости структурные модели можно разделить на три вида: точно идентифицируемые, сверхидентифицируемые и неидентифицируемые (рис. 9.1).

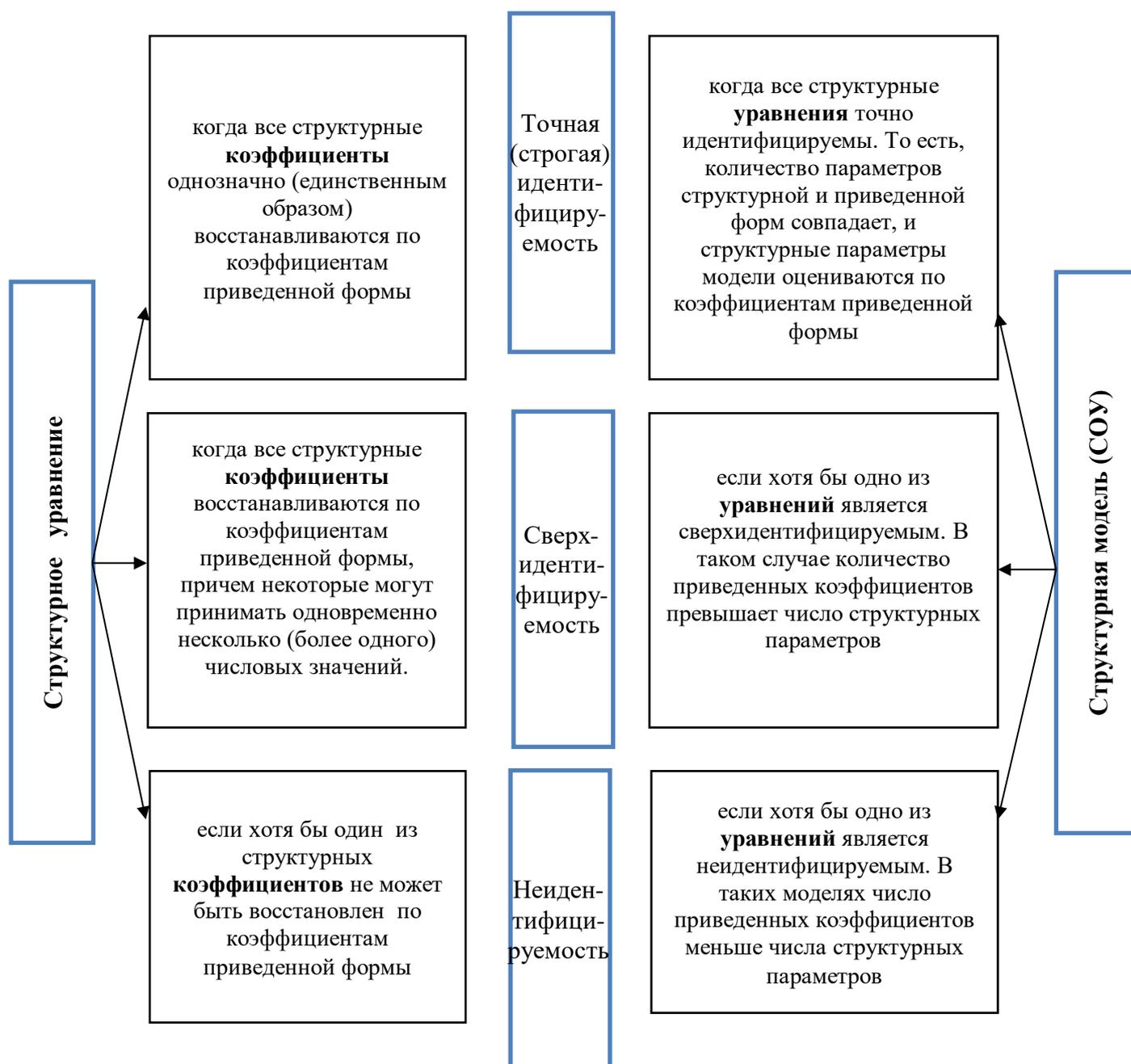


Рисунок 9.1- Идентифицируемость структурных моделей

Проблема идентифицируемости крайне важна, в первую очередь, для решения проблемы *идентификации эконометрической модели*, (т. е. выбора и реализации методов статистического оценивания параметров модели).

Рассмотрим *условия идентифицируемости уравнений системы*

Обозначим общее число predetermined переменных системы через  $k$ , количество predetermined переменных в  $i$ -м уравнении структурной формы через  $k_i$ , число эндогенных переменных модели через  $m$ , а количество эндогенных переменных в  $i$ -м структурном уравнении через  $m_i$ .

С учетом введенной системы обозначений сформулируем *необходимое* (но недостаточное) *условие идентифицируемости*, которое называется «*правило порядка, или счетное правило*»:

**Для идентифицируемости  $i$ -го уравнения число predetermined переменных, отсутствующих в нем, должно быть не меньше числа включенных в него эндогенных переменных без единицы:**

$$k - k_i \geq m_i - 1, \quad \text{или} \quad k_i^* \geq m_i - 1, \quad (9.9)$$

где  $k_i^*$  - количество predetermined переменных, исключенных из  $i$ -го уравнения системы.

Если для всех уравнений модели условие (9.9) выполняется как равенство, то система *точно идентифицирована*. Если хотя бы для одного уравнения имеет место строгое неравенство, то система является *сверхидентифицированной*.

При точной идентифицируемости модели для оценивания параметров ее структурной формы можно применить так называемый *косвенный метод наименьших квадратов*; в случае сверхидентифицируемости системы применение косвенного метода наименьших квадратов уже не ведет к однозначной оценке структурной формы, поэтому следует использовать другие методы, например, *двухшаговый* или *трехшаговый метод наименьших квадратов*.

Если левая часть (9.9) меньше правой части, то имеет место *неидентифицируемость*, и все методы оценивания для совместных линейных моделей теряют силу.

Выполнение условия идентифицируемости модели проверяется для каждого структурного уравнения системы.

Рассмотренное счетное правило, как отмечалось ранее, является необходимым условием идентифицируемости структурной формы модели, но не является достаточным. Более обоснованные выводы относительно идентифицируемости можно сделать, используя *ранговое условие*, являющееся необходимым и достаточным и формулирующееся следующим образом:

*Уравнение идентифицируемо, если по отсутствующим в нем переменным (эндогенным и предопределенным) можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой не равен нулю, а ранг матрицы равен числу эндогенных переменных в системе без единицы*

Целесообразность проверки идентифицируемости модели с помощью условия ранга определяется тем, что возможна ситуация, когда для каждого уравнения системы выполнено счетное правило (необходимое условие), но достаточное условие не выполняется (опредетитель матрицы из названных коэффициентов равен нулю).

В эконометрических моделях часто наряду с регрессионными уравнениями, которые должны быть статистически оценены, присутствуют тождества, имеющие коэффициенты при переменных, равные  $\pm 1$ . В этом случае, хотя само тождество и не требует проверки на идентифицируемость (его коэффициенты известны), в проверке идентифицируемости собственно структурных уравнений модели тождества участвуют.

## 9.6 Оценивание параметров структурной формы эконометрической модели

Для оценивания систем одновременных уравнений существует ряд методов. В целом их можно разбить на две группы. К первой группе относятся методы, применяемые к каждому уравнению в отдельности (например, косвенный метод наименьших квадратов, двухшаговый метод наименьших квадратов, метод максимального правдоподобия с ограниченной информацией, или метод наименьшего дисперсионного отношения). Вторая группа содержит методы, предназначенные для оценивания всей системы в целом (трехшаговый метод наименьших квадратов, метод максимального правдоподобия с полной информацией).

**Косвенный метод наименьших квадратов (КМНК)** (рис. 9.2):

1) преобразование структурной формы модели в приведенную

2) оценивание параметров каждого уравнения приведенной формы с помощью обычного МНК

3) определение оценок структурных параметров по найденным на втором этапе коэффициентам приведенных уравнений с использованием соотношения  

$$C = A^{-1}B$$
, или эквивалентного, но более удобного  $AC = B$

Рисунок 9.2 - Косвенный метод наименьших квадратов

Косвенный метод наименьших квадратов (КМНК) используется в случае точной идентифицируемости модели.

### **Двухшаговый метод наименьших квадратов (2МНК)**

Если система уравнений сверхидентифицируема, то КМНК не используется, поскольку с его помощью невозможно получить однозначные оценки всех параметров структурной формы. В этом случае для оценивания параметров модели *можно применить двухшаговый метод наименьших квадратов (2МНК)*.

Известно, что причиной, по которой при использовании для оценивания параметров структурного уравнения методом наименьших квадратов были бы получены смещенные оценки, является корреляция между случайной составляющей объясняющей (эндогенной) переменной  $Y_t$  и возмущением  $\varepsilon_t$ . Следовательно, если удастся очистить каждое наблюдение  $Y_t$  от его случайной составляющей, то можно будет применить МНК.

Но выделить точно случайную составляющую в каждом наблюдении невозможно, однако можно получить ее оценку с помощью остатка для этого наблюдения, определяемого как  $(Y_t - \hat{Y}_t)$ . Если вычесть эту величину из исходных значений наблюдений вместо самой случайной составляющей, то получим

$$Y_t - (Y_t - \hat{Y}_t) = \hat{Y}_t.$$

Двухшаговый метод наименьших квадратов как раз и реализует описанную процедуру применительно к свержидентифицированному уравнению. Двухшаговый метод наименьших квадратов, как следует из названия, включает два шага (рис. 9.3).

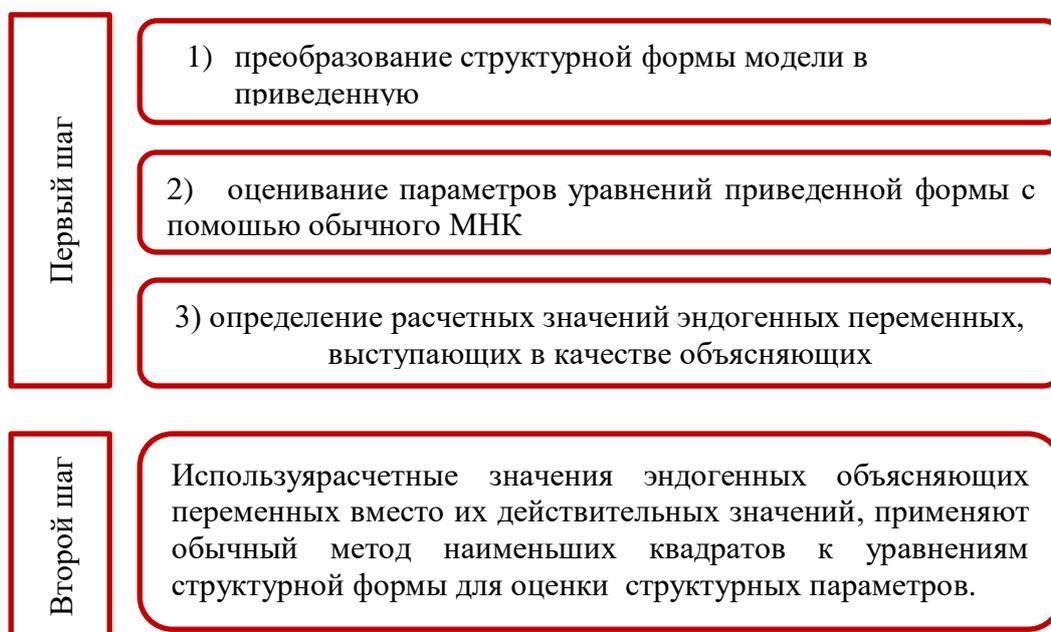


Рисунок 9.3 - Двухшаговый метод наименьших квадратов

Особенности двухшагового метода наименьших квадратов:

- метод можно применять к отдельным уравнениям системы без учета остальных;
- в отличие от косвенного метода наименьших квадратов, который дает несколько разных оценок параметра в сверхидентифицируемом уравнении, метод 2МНК дает лишь одну оценку;
- если для уравнения выполнено ранговое условие идентифицируемости и порядковое условие выполнено со знаком равенства (точная идентифицируемость), то 2МНК-оценка совпадет с оценкой, полученной косвенным методом наименьших квадратов;
- 2МНК-оценка совпадет с оценкой, полученной методом инструментальных переменных, когда в качестве последних используются  $\hat{Y}_1$  и  $X_1$ .

Следует отметить, что в большинстве эконометрических компьютерных пакетов для оценивания систем одновременных уравнений реализован именно 2МНК.

**Трехшаговый метод наименьших квадратов (3МНК)** применяется в случае нарушения предпосылки классической регрессии о гомоскедастичности и (или) об отсутствии автокорреляции возмущений.

Первые два шага в 3МНК – те же, что в 2МНК.

Третий шаг состоит в следующем:

- на основе найденных в 2МНК оценок структурных параметров находят расчетные значения эндогенных объясняемых переменных;
- определяют разности между фактическими и расчетными значениями (остатки);
- по этим остаткам находят оценку ковариационной матрицы случайных составляющих;
- с помощью ковариационной матрицы находят 3МНК-оценки параметров структурной формы.

### Вопросы для самоконтроля

1. Каковы основные причины использования систем одновременных уравнений.
2. Какие трудности возникают при применении МНК, если уравнение содержит случайные объясняющие переменные?
3. Какие переменные системы уравнений называются экзогенными, эндогенными и предопределенными?
4. Какую форму системы уравнений называют структурной формой, а какую приведенной формой системы? В чем состоит основное различие между структурными уравнениями системы и уравнениями в приведенной форме.
5. Какие структурные коэффициенты модели называются идентифицируемыми, неидентифицируемыми, сверхидентифицируемыми?
6. В чем состоят проблемы идентифицируемости модели и какие условия идентифицируемости вы знаете?
7. Назовите возможные способы построения систем уравнений. Чем они отличаются друг от друга?
8. Когда для получения оценок структурной модели можно использовать косвенный метод наименьших квадратов, а когда - двухшаговый метод наименьших квадратов?
9. В чем состоит суть косвенного метода наименьших квадратов?
10. В чем состоит суть двухшагового метода наименьших квадратов?
11. Почему ДМНК (в отличие от КМНК) является общим методом оценивания параметров систем одновременных уравнений?

## СПИСОК ЛИТЕРАТУРЫ

1. Frisch, R. Editorial / R. Frisch // *Econometrica*. – 1933. – № 1. – P. 2.
2. Айвазян, С. А. Прикладная статистика и основы эконометрики : учеб. пособие для вузов / С. А. Айвазян, В. С. Мхитарян. – Москва : ЮНИТИ, 1998. – 1022 с.
3. Бородич, С. А. Эконометрика : учеб. пособие / С. А. Бородич. – 2-е изд., испр. – Минск : Новое знание, 2006. – 408 с.
4. Методические указания к лабораторным и контрольным работам курса «Эконометрика» [Электронный ресурс] / сост.: Ю. Е. Воскобойников, Т. Н. Воскобойникова. - Электрон. дан. - Новосибирск : [б. и.], 2006. - Режим доступа: <http://window.edu.ru/resource/306/63306/files/econometr.pdf>. - Загл. с экрана.
5. Доугерти, К. Введение в эконометрику : учеб. для вузов : пер. с англ. / К. Доугерти. - 2-е изд. - Москва : ИНФРА-М, 2004. – 432 с.
6. Эконометрия : учеб. пособие / В. И. Жлуктенко [и др.]. – Киев : Изд-во Европ. ун-та, 2005. – 552 с.
7. Методические указания к выполнению контрольных заданий по дисциплине «Эконометрия» / сост.: В. Н. Калинина, В. И. Соловьев. – Москва : [б. и.], 2004. – 33 с.
8. Кейн, Э. Экономическая статистика и эконометрия : введение в количественный экономический анализ / Э. Кейн. – Москва : Статистика, 1977. – 255 с.
9. Кремер, Н. Ш. Эконометрика : учеб. для вузов / Н. Ш. Кремер, Б. А. Путко. – Москва : ЮНИТИ-ДАНА, 2002. – 311 с.
10. Лондар, С. Л. Економетрія засобами MS Excel : навч. посіб. / С. Л. Лондар, Р. В. Юринец. – Київ : Вид-во Європ. ун-ту, 2004. – 242 с.
11. Магнус, Я. Р. Эконометрика : нач. курс : учеб. для вузов / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – 6-е изд., перераб. и доп. – Москва : Дело, 2004. - 576 с.

12. Носко, В. П. Эконометрика : учебник. Кн. 1 / В. П. Носко. – Москва : Дело, 2011. – 672 с.
13. Практикум по эконометрике (+CD) : учеб. пособие / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – Москва : Финансы и статистика, 2007. – 344 с. : ил.
14. Айвазян, С. А. Прикладная статистика. Основы эконометрики : учеб. для вузов : в 2 т. Т. 1. Теория вероятностей и прикладная статистика / С. А. Айвазян, В. С. Мхитарян. – Москва : ЮНИТИ-ДАНА, 2001. – 656 с.
15. Айвазян, С. А. Прикладная статистика. Основы эконометрики : учеб. для вузов : в 2 т. Т. 2. Основы эконометрики / С. А. Айвазян. – Москва : ЮНИТИ-ДАНА, 2001. – 432 с.
16. Ферстер, Э. Методы корреляционного и регрессионного анализа : рук. для экономистов : пер. с нем. / Э. Фёрстер, Б. Рёнц. – Москва : Финансы и статистика, 1983. – 302 с.
17. Эконометрика : учебник / К. В. Балдин [и др.] ; под ред. В. Б. Уткина. – 2-е изд. – Москва : Дашков и К°, 2012. – 564 с.
18. Эконометрика : учебник / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – Москва : Финансы и статистика, 2007. – 576 с. : ил.

ПРИЛОЖЕНИЕ А  
МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

**Таблица А.1 – Значения  $F$  -критерия Фишера при уровне значимости  $\alpha = 0,05$**

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	$\infty$
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
$\infty$	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

**Таблица А.2 – Критические значения  $t$ -критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01 (двухсторонний)**

Число степеней свободы d.f.	$\alpha$			Число степеней свободы d.f.	$\alpha$		
	00,10	0,05	0,01		00,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	$\infty$	1,6449	1,9600	2,5758

**Таблица А.3 – Значения статистик Дарбина – Уотсона при уровне значимости  $\alpha = 0,01$**

n*	$m^* = 1$		$m^* = 2$		$m^* = 3$		$m^* = 4$		$m^* = 5$	
	$DW_L$	$DW_U$								
6	0,390	1,142	-	-	-	-	-	-	-	-
7	0,435	1,036	0,294	1,676	-	-	-	-	-	-
8	0,497	1,003	0,345	1,489	0,229	2,102	-	-	-	-
9	0,554	0,998	0,408	1,389	0,279	1,875	0,183	2,433	-	-
10	0,604	1,001	0,466	1,333	0,340	1,733	0,230	2,193	0,150	2,690
11	0,653	1,010	0,519	1,297	0,396	1,640	0,286	2,030	0,193	2,453
12	0,697	1,023	0,569	1,274	0,449	1,575	0,339	1,913	0,244	2,280
13	0,738	1,038	0,616	1,261	0,499	1,526	0,391	1,826	0,294	2,150
14	0,776	1,054	0,660	1,254	0,547	1,490	0,441	1,757	0,343	2,049
15	0,811	1,070	0,700	1,252	0,591	1,464	0,488	1,704	0,391	1,967
16	0,844	1,086	0,737	1,252	0,633	1,446	0,532	1,663	0,437	1,900
17	0,874	1,102	0,772	1,255	0,672	1,432	0,574	1,630	0,480	1,847
18	0,902	1,118	0,805	1,259	0,708	1,422	0,613	1,604	0,522	1,803
19	0,928	1,132	0,835	1,265	0,742	1,415	0,650	1,584	0,561	1,767
20	0,952	1,147	0,863	1,271	0,773	1,411	0,685	1,567	0,598	1,737
21	0,975	1,161	0,890	1,277	0,803	1,408	0,718	1,554	0,633	1,712
22	0,997	1,174	0,914	1,284	0,831	1,407	0,748	1,543	0,667	1,691
23	1,018	1,187	0,938	1,291	0,858	1,407	0,777	1,534	0,698	1,673
24	1,037	1,199	0,960	1,298	0,882	1,407	0,805	1,528	0,728	1,658
25	1,055	1,211	0,981	1,305	0,906	1,409	0,831	1,523	0,756	1,645
26	1,072	1,222	1,001	1,312	0,928	1,411	0,855	1,518	0,783	1,635
27	1,089	1,233	1,019	1,319	0,949	1,413	0,878	1,515	0,808	1,626
28	1,104	1,244	1,037	1,325	0,969	1,415	0,900	1,513	0,832	1,618
29	1,119	1,254	1,054	1,332	0,988	1,418	0,921	1,512	0,855	1,611

30	1,133	1,263	1,070	1,339	1,006	1,421	0,941	1,511	0,877	1,606
31	1,147	1,273	1,085	1,345	1,023	1,425	0,960	1,510	0,897	1,601
32	1,160	1,282	1,100	1,352	1,040	1,428	0,979	1,510	0,917	1,597
33	1,172	1,291	1,114	1,358	1,055	1,432	0,996	1,510	0,936	1,594
34	1,184	1,299	1,128	1,364	1,070	1,435	1,012	1,511	0,954	1,591
35	1,195	1,307	1,140	1,370	1,085	1,439	1,028	1,512	0,971	1,589
36	1,206	1,315	1,153	1,376	1,098	1,442	1,043	1,513	0,988	1,588
37	1,217	1,323	1,165	1,382	1,112	1,446	1,058	1,514	1,004	1,586
38	1,227	1,330	1,176	1,388	1,124	1,449	1,072	1,515	1,019	1,585
39	1,237	1,337	1,187	1,393	1,137	1,453	1,085	1,517	1,034	1,584
40	1,246	1,344	1,198	1,398	1,148	1,457	1,098	1,518	1,048	1,584
45	1,288	1,376	1,245	1,423	1,201	1,474	1,156	1,528	1,111	1,584
50	1,324	1,403	1,285	1,446	1,245	1,491	1,205	1,538	1,164	1,587
55	1,356	1,427	1,320	1,466	1,284	1,506	1,247	1,548	1,209	1,592
60	1,383	1,449	1,350	1,484	1,317	1,520	1,283	1,558	1,249	1,598
65	1,407	1,468	1,377	1,500	1,346	1,534	1,315	1,568	1,283	1,604
70	1,429	1,485	1,400	1,515	1,372	1,546	1,343	1,578	1,313	1,611
75	1,448	1,501	1,423	1,529	1,395	1,557	1,368	1,587	1,340	1,617
80	1,466	1,515	1,441	1,541	1,416	1,568	1,390	1,595	1,364	1,624
85	1,482	1,528	1,458	1,553	1,435	1,578	1,411	1,603	1,386	1,630
90	1,496	1,540	1,474	1,563	1,452	1,587	1,429	1,611	1,406	1,636
95	1,510	1,552	1,489	1,573	1,468	1,596	1,446	1,618	1,425	1,642
100	1,522	1,562	1,503	1,583	1,482	1,604	1,462	1,625	1,441	1,647
150	1,611	1,637	1,598	1,651	1,584	1,665	1,571	1,679	1,557	1,693
200	1,664	1,684	1,653	1,693	1,643	1,704	1,633	1,715	1,623	1,725
*n – число наблюдений, m – количество независимых переменных										

**Таблица А.4 – Значения статистик Дарбина – Уотсона при уровне значимости  $\alpha = 0,05$**

N*	$m^* = 1$		$m^* = 2$		$m^* = 3$		$m^* = 4$		$m^* = 5$	
	$DW_L$	$DW_U$								
6	0,610	1,400	-	-	-	-	-	-	-	-
7	0,700	1,356	0,467	1,896	-	-	-	-	-	-
8	0,763	1,332	0,559	1,777	0,368	2,287	-	-	-	-
9	0,824	1,320	0,629	1,699	0,455	2,128	0,296	2,588	-	-
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,243	2,822
11	0,927	1,324	0,658	1,604	0,595	1,928	0,444	2,283	0,32	2,645
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,379	2,506
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,445	2,390
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,849	0,752	2,023
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991
21	1,222	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,895	1,920
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902

25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,004	1,861
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,819
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813
34	1,393	1,514	1,333	1,580	1,271	1,652	1,028	1,728	1,144	1,808
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780
150	1,720	1,746	1,706	1,760	1,693	1,774	1,679	1,788	1,665	1,802
200	1,758	1,778	1,748	1,789	1,738	1,799	1,728	1,810	1,718	1,820

\* n – число наблюдений, m – количество независимых переменных

**Таблица А.5 – Критические значения хи-квадрат ( $\chi^2$ ) при уровне значимости 0,01; 0,05; 0,1**

Число степеней свободы	$\alpha$			Число степеней свободы	$\alpha$			Число степеней свободы	$\alpha$		
	0,01	0,05	0,1		0,01	0,05	0,1		0,01	0,05	0,1
1	6,6349	3,8415	2,7055	18	34,805	28,869	25,989	35	57,342	49,802	46,059
2	9,2103	5,9915	4,6052	19	36,191	30,144	27,204	36	58,619	50,998	47,212
3	11,345	7,8147	6,2514	20	37,566	31,41	28,412	37	59,893	52,192	48,363
4	13,277	9,4877	7,7794	21	38,932	32,671	29,615	38	61,162	53,384	49,513
5	15,086	11,07	9,2364	22	40,289	33,924	30,813	39	62,428	54,572	50,66
6	16,812	12,592	10,645	23	41,638	35,172	32,007	40	63,691	55,758	51,805
7	18,475	14,067	12,017	24	42,98	36,415	33,196	41	64,95	56,942	52,949
8	20,09	15,507	13,362	25	44,314	37,652	34,382	42	66,206	58,124	54,09
9	21,666	16,919	14,684	26	45,642	38,885	35,563	43	67,459	59,304	55,23
10	23,209	18,307	15,987	27	46,963	40,113	36,741	44	68,71	60,481	56,369
11	24,725	19,675	17,275	28	48,278	41,337	37,916	45	69,957	61,656	57,505
12	26,217	21,026	18,549	29	49,588	42,557	39,087	46	71,201	62,83	58,641
13	27,688	22,362	19,812	30	50,892	43,773	40,256	47	72,443	64,001	59,774
14	29,141	23,685	21,064	31	52,191	44,985	41,422	48	73,683	65,171	60,907
15	30,578	24,996	22,307	32	53,486	46,194	42,585	49	74,919	66,339	62,038
16	32	26,296	23,542	33	54,776	47,4	43,745	50	76,154	67,505	63,167
17	33,409	27,587	24,769	34	56,061	48,602	44,903				

## Предметный указатель

- Автокорреляция 93
- Автокорреляция возмущений 93
- Верификация модели 10
- Гетероскедастичность 82
- Гомоскедастичность 81
- Двухшаговый метод
  - наименьших квадратов 129
- Доступный обобщенный метод
  - наименьших квадратов 107
- Идентификация модели 10
- Идентифицируемость 125,127
- Интервальное оценивание 33,48
- Корреляция 14,17
- Косвенный метод наименьших
  - квадратов 129
- Коэффициент авторегрессии
  - детерминации 30,47
  - корреляции 29
  - частной корреляции 49,50
  - эластичности 35,49
- Коэффициенты приведенной формы 124
  - структурной формы 122
- Метод взвешенных наименьших квадратов
  - 112
  - наименьших квадратов (МНК) 25,26, 44
  - – двухшаговый 129
  - – косвенный 129
  - – обобщенный 105
- Модель линейной регрессии 24
  - множественная 43
  - парная 24
- Мультиколлинеарность 63,
- Неидентифицируемость 126
- Оценка метода наименьших квадратов
  - параметра 24
- Переменная зависимая
  - независимая 24
  - объясняемая 24
  - объясняющая 24
  - фиктивная 72
  - экзогенная 117
  - эндогенная 117
  - предопределенная 119
- Порядковое условие 127
- Приведенная форма модели 124
- Проблема идентифицируемости 125
- Прогноз точечный 34
  - интервальный 34
- Процедура Дарбина 112
  - Кохрейна-Оркатта 113
- Ранговое условие 128
- Распределение
  - Стьюдента (t-распределение) 31,37
  - Фишера (F-распределение) 32,47
  - $\chi^2$  65
- Регрессия 12,17
  - линейная 21
  - нелинейная 22,27
  - пошаговая 67
- Регрессионная модель линейная 21,
  - – множественная 43
  - обобщенная 102,103
  - парная 24
- Сверхидентифицируемость 126
- Системы одновременных уравнений 18,116,
  - 118
  - взаимозависимых 121
  - независимых 120
- Спецификация модели 10
- Стандартное отклонение (ошибки) 46,47
- Статистика Дарбина-Уотсона 97
  - Фаррара-Глоубера 65
- Статистическая значимость 31
  - коэффициентов регрессии 31
  - уравнения регрессии 32
- Структурная форма модели 122
- Структурные коэффициенты 122
- Теорема Эйткена 105,107
- Тест Бреуша-Пагана 89
  - Бреуша-Годфри 98
  - Глейзера 90
  - Голдфельда-Квандта 86
  - Дарбина-Уотсона 97
  - Парка 88
  - Спирмена 87
  - Уайта 90
  - Чоу 79
- Точная идентифицируемость 126
- Эконометрика 8, 10,
- Эконометрическая модель 17,18