

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
И ПРОДОВОЛЬСТВИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

ГЛАВНОЕ УПРАВЛЕНИЕ ОБРАЗОВАНИЯ, НАУКИ И КАДРОВ

Учреждение образования
«БЕЛОРУССКАЯ ГОСУДАРСТВЕННАЯ
СЕЛЬСКОХОЗЯЙСТВЕННАЯ АКАДЕМИЯ»

Е. В. Карачевская, С. П. Сазонова

ЭКОНОМЕТРИКА (ПРОДВИНУТЫЙ УРОВЕНЬ)

Курс лекций

для магистрантов, обучающихся по специальностям

1-25 80 04 Эконометрика и управление народным хозяйством,

1-25 80 05 Бухгалтерский учет, статистика

Горки
БГСХА
2020

УДК 330.43:519.862.6(075.8)

ББК 65.050я73

К21

*Рекомендовано методической комиссией
экономического факультета 26.06.2018 (протокол № 10)
и Научно-методическим советом БГСХА
28.06.2018 (протокол № 10)*

Авторы:

кандидат экономических наук, доцент *Е. В. Карачевская*;
старший преподаватель *С. П. Сазонова*

Рецензенты:

кандидат экономических наук, доцент *А. В. Грибов*;
главный специалист управления по сельскому хозяйству
и продовольствию Горещкого районного исполнительного комитета
А. А. Курляндчик

Карачевская, Е. В.

К21 Эконометрика (продвинутый уровень) : курс лекций /
Е. В. Карачевская, С. П. Сазонова. – Горки : БГСХА, 2020. – 82 с.
ISBN 978-985-882-042-8.

Приведен теоретический материал лекций для самостоятельного изучения некоторых разделов программы курса. Даны теоретические аспекты рассматриваемых тем.

Для магистрантов, обучающихся по специальностям 1-25 80 04 Эконометрика и управление народным хозяйством, 1-25 80 05 Бухгалтерский учет, статистика.

УДК 330.43:519.862.6(075.8)

ББК 65.050я73

ISBN 978-985-882-042-8

© УО «Белорусская государственная
сельскохозяйственная академия», 2020

ВВЕДЕНИЕ

Цель данного курса – научить студентов использовать основные методы эконометрики, необходимые для проверки предлагаемых и выявления новых эмпирических зависимостей, а также дать представление о современной инструментари эконометрического моделирования, ознакомить их с практическим применением методов эконометрики при проведении научных и прикладных экономических исследований на основе экономической теории и реальных статистических данных, с использованием современных прикладных программ и компьютерных технологий.

Задачи курса:

- изучить принципы количественного анализа реальных экономических процессов и явлений во времени и в пространстве;
- получить знания по эмпирическому выводу экономических зависимостей, закономерностей и законов, действующих в настоящее время;
- научиться строить и использовать эконометрические модели, а также оценивать их параметры для объяснения поведения исследуемых экономических явлений;
- проверять выдвигаемые гипотезы о свойствах экономических показателей и формах их связи;
- научиться оценивать и использовать результаты экономического анализа для прогноза и принятия обоснованных экономических решений.

1. ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ (ПРОДВИНУТЫЙ УРОВЕНЬ)

В настоящее время деятельность в любой сфере области экономики требует от специалиста применения современных методов работы, знаний достижений мировой экономической мысли. Большинство новых методов основано на эконометрических моделях, концепциях, приемах. Без глубоких знаний эконометрики научиться их использовать невозможно.

Эконометрика, опираясь на теорию вероятностей и математическую статистику, должна вооружить экономиста, бухгалтера, менеджера и других специалистов современным эконометрическим инструментарием. Не владея им, специалист является неконкурентоспособным на рынке труда.

Успешная работа современного экономиста в любой области экономики тесным образом связана с использованием математических методов и средств вычислительной техники. При решении задач из различных областей человеческой деятельности часто приходится использовать методы, основанные на эконометрических моделях. Эконометрика – одна из базовых дисциплин экономического образования во всем мире, данный предмет только начал входить в учебные планы обучения будущих экономистов, так как прежде в СССР в условиях централизованной плановой экономики эконометрика была просто не нужна.

Эконометрика (как учебный предмет) призвана вооружить экономиста, менеджера, инженера современным эконометрическим инструментарием, разработанным за последние 50–70 лет. Не владея эконометрикой, отечественный специалист – менеджер и инженер – оказывается неконкурентоспособным по сравнению с зарубежным. Во многих странах мира – Японии и США, Франции и Швейцарии, Перу и Ботсване и др. – статистическим методам обучают в средней школе, ЮНЕСКО постоянно проводят конференции по вопросам такого обучения. В СССР и СЭВ, а теперь и в России, игнорируют этот предмет в средней школе и лишь слегка затрагивают его в высшей. Результат на рынке труда очевиден – снижение конкурентоспособности специалистов.

Эконометрика – быстроразвивающаяся отрасль науки, цель которой состоит в том, чтобы придать количественные меры экономическим отношениям.

Термин «эконометрика» был впервые введен бухгалтером П. Цемпой (Австро-Венгрия, 1910). Цемпа считал, что если к данным бухгалтерского учета применить методы алгебры и геометрии, то будет получено новое, более глубокое представление о результатах хозяйственной деятельности. Это употребление термина, как и сама концепция, не прижилось, но название «эконометрика» оказалось весьма удачным для определения нового направления в экономической науке, которое выделилось в 1930 г.

Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика» (от греч. «метрон»). Таким образом, сам термин подчеркивает специфику, содержание эконометрики как науки: количественное выражение тех связей и соотношений, которые раскрыты и обоснованы экономической теорией. И. Шумпетер (1883–1950), один из первых сторонников выделения этой новой дисциплины, полагал, что в соответствии со своим назначением эта дисциплина должна называться «эконометрика». Советский ученый А. Л. Вайнштейн (1892–1970) считал, что название настоящей науки основывается на греческом слове «метрия» (геометрия, планиметрия и т. д.), соответственно по аналогии – эконометрия. Однако в мировой науке общепринятым стал термин «эконометрика». В любом случае, какой бы мы термин ни выбрали, эконометрика является наукой об измерении и анализе экономических явлений.

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эта наука возникла в результате взаимодействия и объединения в особый «сплав» трех компонент: экономической теории, статистических и математических методов. Впоследствии к ним присоединилось развитие вычислительной техники как условие развития эконометрики.

В журнале «Эконометрика», основанном в 1933 г. Р. Фришем (1895–1973), он дал 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 трех отправных точек – статистика, экономическая теория и математика – необходимое, но недостаточное условие для понимания количественных соотношений в современной экономической жизни. Это единство всех трех составляющих. И это единство образует эконометрику».

Таким образом, эконометрика – это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов.

Структура эконометрики.

В эконометрике, как дисциплине на стыке экономики (включая менеджмент) и статистического анализа, естественно выделить три вида научной и прикладной деятельности (по степени специфичности методов, сопряженной с погруженностью в конкретные проблемы):

а) разработка и исследование эконометрических методов (методов прикладной статистики) с учетом специфики экономических данных;

б) разработка и исследование эконометрических моделей в соответствии с конкретными потребностями экономической науки и практики;

в) применение эконометрических методов и моделей для статистического анализа конкретных экономических данных.

Кратко рассмотрим три только что выделенных вида научной и прикладной деятельности. По мере движения от «а» к «в» сужается широта области применения конкретного эконометрического метода, но при этом повышается его значение для анализа конкретной экономической ситуации. Если работам вида «а» соответствуют научные результаты, значимость которых оценивается по общеэконометрическим критериям, то для работ вида «в» основное – успешное решение задач конкретной области экономики. Работы вида «б» занимают промежуточное положение, поскольку, с одной стороны, теоретическое изучение эконометрических моделей может быть весьма сложным и математизированным, с другой – результаты представляют интерес не для всей экономической науки, а лишь для некоторого направления в ней.

Ситуация с внедрением современных статистических (эконометрических) методов на предприятиях и в организациях различных отраслей народного хозяйства противоречива. К сожалению, при развале отечественной промышленности в 1990-е гг. больше всего пострадали структуры, наиболее нуждающиеся в эконометрических методах, – службы качества, надежности, центральные заводские лаборатории и др. Однако толчок к развитию получили службы маркетинга и сбыта, сертификации, прогнозирования, инноваций и инвестиций, которым также полезны различные эконометрические методы, в частности, методы экспертных оценок.

Предмет эконометрики.

К основным задачам эконометрики относятся:

- Построение экономических моделей в математической форме, удобной для эмпирического анализа (проблема спецификации).
- Определение параметров уравнения (этап параметризации).
- Проверка качества найденных параметров модели и самой модели в целом (верификация).
- Использование построенных моделей для объяснения поведения исследуемых экономических показателей, прогнозирования и предсказания, а также для осмысленного проведения экономической политики.

Основным **предметом** исследования эконометрики являются массовые экономические явления и процессы.

Эконометрика через математические и статистические методы анализирует экономические закономерности, доказанные экономической теорией.

Принципы эконометрики:

- принцип правильной постановки проблемы;
- системной направленности эконометрических расчетов;
- учета рыночной неопределенности;
- улучшения имеющихся альтернатив и поиска новых.

Основные эконометрические методы.

1. Сводка и группировка информации.

Статистическая сводка – это научно организованная обработка материалов наблюдения, включающая в себя систематизацию, группировку данных, составление таблиц, подсчет итогов, расчет производных показателей (средних, относительных величин).

Статистическая группировка – это процесс образования однородных групп на основе расчленения статистической совокупности на части или объединения изучаемых единиц в частные совокупности по существенным для них признакам.

2. Вариационный и дисперсионный анализ.

Дисперсия признака – это средний квадрат отклонений вариантов от их средней величины. В эконометрических расчетах, как правило, используют общую, межгрупповую и внутригрупповую дисперсии. При этом общая дисперсия характеризует вариацию признака в статистической совокупности в результате влияния всех факторов. Межгрупповая дисперсия показывает размер отклонения групповых средних от общей средней, т. е. характеризует влияние фактора, положенного в основание группировки. Внутригрупповая (остаточная) дисперсия характеризует вариацию признака в середине каждой группы ста-

тистической группировки. В эконометрических расчетах используется среднее квадратическое отклонение – обобщающая характеристика размеров вариации признака в совокупности. Оно равно корню квадратному из дисперсии. Для осуществления сравнений колеблемости одного и того же признака в нескольких совокупностях используется относительный показатель вариации – коэффициент вариации.

3. Регрессионный и корреляционный анализ.

Применение метода наименьших квадратов (МНК) позволяет получить достаточно точные теоретические значения модели однофакторной регрессии и соответственно ее графическое изображение (термин «регрессия» – движение назад, возвращение в прежнее состояние, – был введен Фрэнсисом Галтоном в конце XIX в. при анализе зависимости между ростом родителей и ростом детей; в любом случае средний рост детей – и у низких, и у высоких родителей – стремится (возвращается) к среднему росту людей в данном регионе).

4. Статистические уравнения зависимости.

5. Статистические индексы и др.

Статистические индексы могут быть использованы в качестве меры изменения количества независимо от изменения качественного признака (цены, себестоимости, производительности труда и т. п.), а также для характеристики качественного признака независимо от изменения количества (объема продукции в натуральном выражении, численности работников и т. п.).

2. ОБЩАЯ ЛИНЕЙНАЯ СТАТИСТИЧЕСКАЯ И ЕЕ ПОСТРОЕНИЕ С ПОМОЩЬЮ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

2.1. Общий вид эконометрических моделей. Основные понятия и этапы построения.

2.2. Классификация эконометрических моделей.

2.3. Парная и множественная регрессия и корреляция.

2.4. Модели множественной линейной регрессии.

2.1. Общий вид эконометрических моделей. Основные понятия и этапы построения

Функция $f(x)$ называется *функцией регрессии* у на x , если она описывает поведение условного среднего значения зависимой перемен-

ной y (при условии, что значения объясняющей(их) переменной(ых) зафиксированы).

Зависимость такого типа выражается соотношением

$$M(y / x) = f(x), \quad (2.1)$$

где x – независимая (объясняющая) переменная (регрессор);

y – зависимая (объясняемая) переменная.

При рассмотрении зависимости двух СВ говорят о *парной регрессии*. Символ $M(y / x)$ означает условное математическое ожидание (математическое ожидание y при заданном значении x).

Парная регрессия – это уравнение, описывающее корреляционную связь между парой переменных: зависимой переменной (результатом) y и независимой переменной (фактором) x : $y = f(x)$. Функция может быть как линейной, так и нелинейной, например

$$\hat{y} = \alpha + \frac{\beta}{x} + \varepsilon, \text{ или } y = \alpha + \beta x + \varepsilon \text{ и т. д.}$$

Мы будем рассматривать парную регрессию, описывающую линейную связь между двумя переменными, которая представлена в следующем виде:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = \overline{1, N}, \quad (2.2)$$

где y_i – i -е значение зависимой переменной y ;

x_i – i -е значение независимой переменной x ;

α, β – генеральные параметры парной линейной регрессии;

N – объем генеральной совокупности.

Это уравнение можно использовать для изучения зависимости потребления (y) от уровня доходов (x), инвестиций (y) от процентной ставки (x) и для многих других задач. Уравнение парной регрессии в виде (1.2) относится к генеральной совокупности. Практически регрессия строится по данным выборки и записывается в виде

$$y_i = a + bx_i + e_i, \quad i = \overline{1, n},$$

где n – объем выборки;

a, b – выборочные оценки параметров парной линейной регрессии.

Параметр a называют свободным членом регрессии; параметр b – ко-

ээффициент регрессии, который измеряет, на сколько единиц в среднем изменится y при изменении x на одну единицу.

Итак, задача состоит в том, чтобы по данным выборки оценить неизвестные параметры линейной парной регрессии: параметр α – пересечение (*intercept*) и параметр β – наклон линии регрессии (*slope*) (рис. 2.1).

Как показано на графике, параметр α соответствует отрезку прямой, отсекаемому линией регрессии при пересечении ею оси ординат. Параметр β определяет наклон линии регрессии к оси абсцисс.

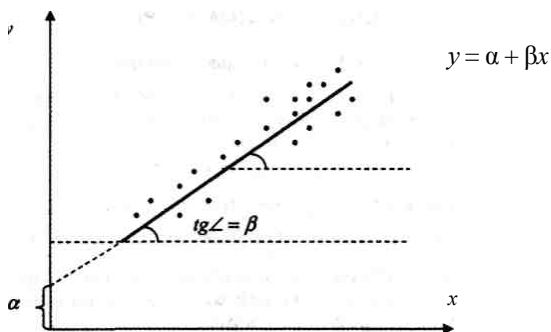


Рис. 2.1. Линия регрессии

Зависимость нескольких переменных, выражаемую функцией, называют множественной регрессией:

$$M(y / x_1, x_2, \dots, x_m) = y = f(x_1, x_2, \dots, x_p, e), \quad (2.3)$$

где y – зависимая переменная (результат);

x_1, x_2, \dots, x_p – независимые переменные (факторы);

e – случайный остаток;

f – некая математическая функция.

В качестве функций множественной регрессии часто выбирают наиболее простые – линейную, показательную и степенную:

$y = a + b_1 x_1 + \dots + b_p x_p + e$ – линейная функция;

$y = a x_1^{b_1} \cdot \dots \cdot x_p^{b_p} e$ – степенная функция;

$y = a b_1^{x_1} \cdot \dots \cdot b_p^{x_p} e$ – показательная функция, где a, b_p – параметры функций.

Эти функции могут быть использованы и при формировании смешанных моделей.

При проведении эконометрического анализа предполагается, что наблюдения, на основе которых он проводится, были получены по однородной совокупности единиц. То есть механизм воздействия факторов на результат должен быть примерно одинаков на разных единицах совокупности. Для обеспечения статистической достоверности модели количество наблюдений должно быть в 8–10 раз больше количества параметров, не считая свободного члена.

Реальные значения зависимой переменной не всегда совпадают с ее условными математическими ожиданиями и могут быть различными при одном и том же значении объясняющей переменной (наборе объясняющих переменных), поэтому фактическая зависимость должна быть дополнена некоторым слагаемым, которым является СВ.

Рассмотрим основные причины обязательного присутствия в регрессионных моделях случайного фактора (отклонения) ε .

1. *Невключение в модель всех объясняющих переменных.* Любая регрессионная (в частности, эконометрическая) модель является упрощением реальной ситуации. Последняя всегда представляет собой сложнейшее переплетение различных факторов, многие из которых в модели не учитываются, что порождает отклонение реальных значений зависимой переменной от ее модельных значений.

2. *Неправильный выбор функциональной формы модели.* Из-за слабой изученности исследуемого процесса либо из-за его переменчивости может быть неверно подобрана моделирующая его функция. Это, безусловно, скажется на отклонении модели от реальности, что отразится на величине случайного члена. Кроме того, неверным может быть подбор объясняющих переменных.

3. *Агрегирование переменных.* Во многих моделях рассматриваются зависимости между факторами, которые сами представляют сложную комбинацию других, более простых переменных. Это может оказаться причиной отклонения реальных значений от модельных.

4. *Ошибки измерений.* Какой бы качественной ни была модель, ошибки измерений переменных отразятся на несоответствии модельных значений эмпирическим данным, что также отразится на величине случайного члена.

5. *Ограниченность статистических данных.* Нередко строятся модели, выражаемые непрерывными функциями. Но для этого используется набор данных, имеющих дискретную структуру.

Это несоответствие находит свое выражение в случайном отклонении.

6. *Непредсказуемость человеческого фактора.* Эта причина может «испортить» самую качественную модель, так как невозможно спрогнозировать поведение каждого индивидуума.

Решение задачи построения качественного уравнения регрессии, соответствующего эмпирическим данным и целям исследования, является достаточно сложным и многоступенчатым процессом. Его можно разбить на пять этапов:

- **первый** – теоретический, в ходе которого формулируется цель исследования, определяется круг участвующих в модели экономических характеристик, создается априорное описание формализованных связей между ними;

- **второй** – информационный, когда осуществляется поиск требуемых данных, осуществляются необходимые пересчеты, используются пространственные и временные данные;

- **третий** – спецификация модели, когда устанавливаются связи и соотношения, т. е. выбор формулы связи переменных (и самих переменных, включаемых в уравнение). В случае парной регрессии выбор формулы обычно осуществляется по графическому изображению реальных статистических данных в виде точек в декартовой системе координат, которое называется *корреляционным полем (диаграммой рассеяния)*;

- **четвертый** – идентификация модели, т. е. выявление условий корректного оценивания параметров модели на основе соотношения количества переменных и связей между ними; оценка параметров модели;

- **пятый** – верификация модели, т. е. проверка адекватности модели, делается вывод о том, какова точность расчетов на основе модели, получаемых прогнозных оценок; производится анализ остатков (случайных компонент).

В этой связи при обосновании результативного и факторных показателей необходимо руководствоваться следующими положениями:

1. *Результативный показатель в цепочке причинно-следственных связей всегда находится на более высоком уровне, который определяют на основе логических рассуждений, а также знаний о том, какие из рассматриваемых показателей являются первичными и вторичными.*

Например, y_x – себестоимость, x – урожайность или факторный показатель – урожайность (x), результативный – себестоимость (y_x).

2. В корреляционную модель следует включать факторы, которые оказывают непосредственное влияние на результативный показатель.

Например, урожайность ячменя является результативным показателем (y). Доля или процент посевов ячменя в зерновом клине – это фактор (x). Механизм влияния данного фактора на урожайность отсутствует, хотя известно, что доля посевов обладает более высокой урожайностью, чем другие культуры. Такие факторы не следует включать в корреляционную модель. Размер доли посевов не влияет на формирование показателя урожайности.

3. В корреляционную модель включают факторы, которые логически определяют содержание результативного показателя.

Например, стоимость валовой продукции y_x – результативный показатель. При этом факторы: x_1 – основные производственные фонды, у. е.; x_2 – оборотные фонды, у. е.; x_3 – сельскохозяйственные угодья, га; x_4 – балл 1 га сельскохозяйственных угодий; x_5 – среднегодовые работники, чел.

4. Если результативный показатель является синтетическим (сложным), то и факторные показатели должны быть такими же.

Например, стоимость валовой продукции – это показатель синтетический, который характеризует хозяйство в целом. В свою очередь, остальные показатели также являются синтетическими и сложными. Для этого вводим факторы: x_6 – стоимость зданий; x_7 – стоимость силовых машин и оборудования; x_8 – остальные фонды. При этом $x_6 + x_7 + x_8 = x_1$.

5. Если результативный показатель является относительным, то и факторные показатели должны быть такими же (по возможности).

Например, результативный показатель – это уровень производства, который равен отношению стоимости валовой продукции к площади сельскохозяйственных угодий (из расчета на 100 га). Значит, все остальные показатели должны быть взяты в расчете на 100 га сельскохозяйственных угодий, за исключением тех, содержание которых от количества гектаров не меняется (например, балл сельскохозяйственных угодий).

6. Если исследования показывают, что увеличение какого-то показателя сверх определенного уровня предполагает получение дополнительного эффекта, то этот показатель может быть учтен дважды, что вытекает из закона превращения количества в качество.

Например, установлено, что если стоимость кормов в издержках превышает 28 %, то эффективность оборотных фондов возрастает. Чтобы посчитать эффективность превышения фондов сверх 28 %, вводим дополнительный фактор (стоимость кормов сверх 28 % в стоимости оборотных фондов), что позволит определить его влияние на резуль- тативный показатель.

Вид эконометрической модели определяется следующими спосо- бами:

- 1) логического анализа;
- 2) графиков;
- 3) аналитического приема.

1. Способ логического анализа основан на логических рассуждени- ях о взаимосвязи зависимой переменной и независимой.

2. Графический метод основан на построении точечных графиков (поле корреляции).

В случае если речь идет об однофакторной эконометрической мо- дели, то вывод о характере связи y и x делают на основании одного графика. Для этого строим корреляционное поле и по расположению в нем точек находим преобладающую тенденцию (рис. 2.2, 2.3).

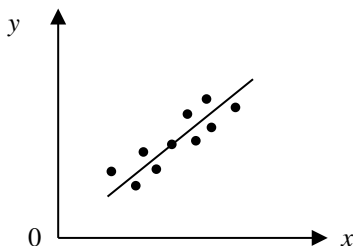


Рис. 2.2. Корреляционное поле линейной зависимости

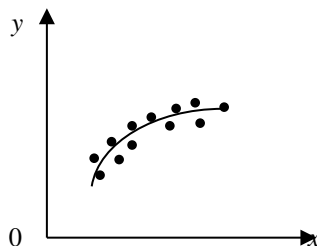


Рис. 2.3. Корреляционное поле нелинейной зависимости

Из взаимосвязи y и x , показанной на рис. 2.2, следует, что эконо- метрическая модель имеет вид $y_x = a_0 + a_1x$. В случае, показанном на рис. 2.3, характер связи имеет такой вид: $y_x = a_0x^{a_1}$.

Если корреляционное поле отличается неопределенностью, то влия- ние фактора x_2 учитываем как линейное, т. е. a_2x_2 .

3. Вид эконометрической модели определяется с помощью ПК (на практических занятиях).

Таким образом, на основании графиков, интуиции, опыта с помощью ПК можно определить вид корреляционной модели.

2.2. Классификация эконометрических моделей

Выделяют три основных класса эконометрических моделей.

1. Модель временных рядов.

Модель представляет собой зависимость результативного признака от переменной времени или переменных, относящихся к другим моментам времени.

К моделям временных рядов, в которых результативный признак зависит от времени, относятся:

- **модель тренда** (модель зависимости результативного признака от трендовой компоненты):

$$y(t) = T(t) + \varepsilon_t,$$

где $y(t)$ – результативный признак;

$T(t)$ – временной тренд заданного параметрического вида (например, линейный);

ε_t – случайная (стохастическая) компонента;

- **модель сезонности** (модель зависимости результативного признака от сезонной компоненты):

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

где $S(t)$ – периодическая (сезонная) компонента;

ε_t – случайная (стохастическая) компонента;

- **модель тренда и сезонности:**

$$y(t) = T(t) + S(t) + \varepsilon_t \text{ (аддитивная) или}$$
$$y(t) = T(t)S(t) + \varepsilon_t \text{ (мультипликативная),}$$

где $T(t)$ – временной тренд заданного параметрического вида;

$S(t)$ – периодическая (сезонная) компонента;

ε_t – случайная (стохастическая) компонента.

К моделям временных рядов, в которых результативный признак зависит от переменных, датированных другими моментами времени, относятся:

- **модели с распределенным лагом**, которые объясняют вариацию результативного признака в зависимости от предыдущих значений факторных переменных;

- **модели авторегрессии**, которые объясняют вариацию результативного признака в зависимости от предыдущих значений результативных переменных;

- **модели ожидания**, объясняющие вариацию результативного признака в зависимости от будущих значений факторных или результативных переменных.

Модели временных рядов подразделяются на **модели, построенные по стационарным и нестационарным временным рядам**.

Стационарные временные ряды характеризуются постоянными во времени средней, дисперсией и автокорреляцией, т. е. данный временной ряд не содержит трендовую и сезонную компоненты.

Если временной ряд не отвечает перечисленным условиям, то он является **нестационарным** (т. е. содержит трендовую и сезонную компоненты).

2. Регрессионные модели с одним уравнением.

В таких моделях зависимая (объясняемая) переменная y представляется в виде функции

$$y = f(x, \beta) = f(x_1, x_2, \dots, x_k, \beta_1, \beta_2, \dots, \beta_p),$$

где x_1, x_2, \dots, x_k – независимые (объясняющие) переменные;

$\beta_1, \beta_2, \dots, \beta_p$ – параметры.

Регрессионные модели подразделяются:

- на **парные** (с одним факторным признаком) ($y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$);

- **множественные (с несколькими факторными признаками)** ($y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon_i$).

В зависимости от вида функции $f(x, \beta)$ модели подразделяются:

- на **линейные** ($y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon_i$);

- **нелинейные** ($y_i = \beta_0 + \beta_1 \sqrt{x_1} + \beta_2 \sqrt{x_2} + \dots + \beta_n \log x_n + \varepsilon_i$).

3. Системы одновременных уравнений.

Данные модели описываются системами взаимозависимых регрессионных уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может включать в себя не только факторные переменные, но и результативные переменные из других уравнений системы.

Для тождеств характерно то, что их вид и значения параметров известны. Регрессионные уравнения, из которых состоит система, называются поведенческими уравнениями. В поведенческих уравнениях значения параметров являются неизвестными и подлежат оцениванию.

Примером может служить модель спроса и предложения, приведенная ниже. Системы одновременных уравнений требуют относительно более сложный математический аппарат. Они могут использоваться для моделей страновой экономики и др.

Пример. Модель спроса и предложения. Пусть Q_t^D – спрос на товар в момент времени t , Q_t^S – предложение товара в момент времени t , P_t – цена товара в момент времени t , Y_t – доход в момент времени t . Составим следующую систему уравнений «спрос-предложение»:

$$Q_t^S = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_t \text{ (предложение);}$$

$$Q_t^D = \beta_1 + \beta_2 P_t + \beta_3 Y_{t-1} + u_t \text{ (спрос);}$$

$$Q_t^S = Q_t^D \text{ (равновесие).}$$

Цена товара P_t и спрос на товар $Q_t = Q_t^S = Q_t^D$ определяются из уравнений модели, т. е. являются эндогенными переменными. Предопределенными переменными в данной модели являются доход Y_t и значение цены товара в предыдущий момент времени P_{t-1} .

Классификация видов эконометрических переменных и типов данных.

В эконометрических исследованиях, как правило, используются два типа выборочных данных:

- 1) пространственные (crosssectional data);
- 2) временные (timeseries data).

Под *пространственными данными* понимается совокупность экономической информации, относящейся к разным объектам, полученной за один и тот же период или момент времени. Пространственные данные представляют собой выборочную совокупность из некоторой генеральной совокупности. В качестве примера пространственных данных можно привести совокупность различной информации по какому-либо предприятию (численность работников, объем производства, размер основных фондов), об объемах потребления продукции определенного вида и т. д.

Под *временными данными* понимается совокупность экономической информации, характеризующей один и тот же объект, но за разные периоды времени. По аналогии с пространственной выборкой отдельно взятый временной ряд можно считать выборкой из бесконечного ряда значений показателей во времени. В качестве примера временных данных можно привести данные о динамике индекса потребитель-

ских цен, ежедневные обменные курсы валют. Временная информация естественным образом упорядочена во времени в отличие от пространственных данных.

Экономические переменные, участвующие в любой эконометрической модели, подразделяются на четыре вида:

1) **экзогенные** (независимые) – переменные, значения которых задаются извне. В определенной степени данные переменные являются управляемыми (x);

2) **эндогенные** (зависимые) – переменные, значения которых определяются внутри модели, или взаимозависимые (y);

3) **лаговые** – экзогенные или эндогенные переменные в эконометрической модели, относящиеся к предыдущим моментам времени и находящиеся в уравнении с переменными, относящимися к текущему моменту времени. Например, x_{i-1} – лаговая экзогенная переменная, y_{i-1} – лаговая эндогенная переменная;

4) **предопределенные** (объясняющие переменные) – лаговые (x_{i-1}) и текущие (x) экзогенные переменные, а также лаговые эндогенные переменные (y_{i-1}). Любая эконометрическая модель предназначена для объяснения значений одной или нескольких текущих эндогенных переменных в зависимости от значений предопределенных переменных.

2.3. Парная и множественная регрессия и корреляция

Рассмотрим общую схему процедуры оценки параметров линейной эконометрической модели на основе МНК более подробно. Такая модель в общем виде была представлена уравнением (2.1):

$$y_x = \alpha_0 + \alpha_1 x_t + \dots + \alpha_n x_m + \varepsilon. \quad (2.4)$$

Исходными данными при оценке параметров $\alpha_0, \alpha_1, \dots, \alpha_m$ являются измеренные (наблюдаемые) значения зависимой переменной.

Свое название МНК получил исходя из смыслового содержания критерия, которому должны удовлетворять полученные на его основе оценки параметров эконометрической модели: **сумма квадратов значений фактической ошибки модели должна быть минимальной.**

Иными словами, найденные с помощью МНК оценки $\alpha_0, \alpha_1, \dots, \alpha_m$ обеспечивают минимум следующей квадратичной формы на множестве ее других комбинаций значений таких оценок:

$$s^2 = \sum_{n=1}^N e^2 = \sum_{n=1}^N (y_x - \alpha_0 - \alpha_1 x_1 - \dots - \alpha_m x_m)^2; \quad (2.5)$$

$$\sum (y_x - y_i)^2 \rightarrow \min,$$

где e – значение фактической ошибки модели, полученное после подстановки в выражение (2.4) вместо неизвестных истинных значений параметров $\alpha_0, \alpha_1, \dots, \alpha_m$ их оценок a_0, a_1, \dots, a_m .

Оптимальные по данному критерию значения оценок в этом случае могут быть найдены решением системы так называемых нормальных уравнений, вытекающей из условия равенства нулю частных производных функции $s^2 = (\alpha_0, \alpha_1, \dots, \alpha_n)$ по своим параметрам в точке минимума.

Характеристики эконометрической модели.

Отметим основные критерии оценки «качества» эконометрических моделей.

1. Коэффициенты тесноты связи:

а) для линейной однофакторной ЭМ рассчитывают *коэффициент парной корреляции*:

$$r_{xy} = \frac{\overline{xy} - \overline{x}\overline{y}}{\delta_y \delta_x}, \quad (2.6)$$

где \overline{xy} – среднее произведения;

$\overline{x}\overline{y}$ – произведение средних;

$\delta_x \delta_y$ – произведение среднеквадратических отклонений соответственно факторного и результативного показателей.

Коэффициент парной корреляции изменяется в пределах $-1 \leq r \leq 1$ и показывает силу и направление связи между результативным и факторным показателями;

б) если ЭМ многофакторная линейная, определяют *коэффициент множественной корреляции* R .

Общая формула коэффициента множественной корреляции (корреляционного отношения):

$$R(\eta) = \sqrt{1 - \frac{\sum (y_x - \bar{y})^2}{\sum (y_i - \bar{y})^2}}, \quad (2.7)$$

где y_x – расчетное значение результативного показателя;

y_i – фактическое значение результативного показателя;

\bar{y} – среднее значение фактического результативного показателя.

Значения коэффициента множественной корреляции и корреляционного отношения изменяются в пределах $0 \leq R(\eta) \leq 1$ и отражают силу влияния учтенных в модели факторных признаков на результативный. Чем ближе показатель к единице, тем связь сильнее. Если $0 \leq R(\eta) \leq 0,3$, считают, что связь между результативным и включенными в модель факторными показателями слабая, если $0,3 < R(\eta) \leq 0,7$ – средняя, если $0,7 \leq R(\eta) < 1,0$ – сильная;

в) если ЭМ нелинейная, определяют *корреляционное отношение* (η):

$-1 \leq \eta \leq 1$, $0 \leq \eta \leq 1$ (показывает то же и рассчитывается так же, как r , R).

2. Так как коэффициенты тесноты связи величины вероятностные, их необходимо проверить на существенность, т. е. определить **коэффициент существенности коэффициента множественной (парной) корреляции или корреляционного отношения**:

$$t_{R(\eta)} = \frac{R(\eta)}{\mu_{R(\eta)}}, \quad t_r = \frac{r}{\mu_r}, \quad (2.8)$$

где $\mu_r \cdot \mu_{R(\eta)}$ – соответственно ошибка коэффициента парной, множественной корреляции (корреляционного отношения).

Ошибка коэффициента парной, множественной корреляции (корреляционного отношения) находится по формуле

$$\mu_{R(\eta)} = \frac{1 - R^2(\eta^2)}{\sqrt{n - k - 1}}, \quad \mu_r = \frac{1 - r^2}{\sqrt{n - 1}}, \quad (2.9)$$

где k – количество факторов модели;

n – количество опытов.

Если коэффициент существенности коэффициента парной, множественной корреляции (корреляционного отношения) больше его табличного значения (2,48), то коэффициент парной, множественной корреляции (корреляционного отношения) существенен, и модель, для которой он рассчитан, можно применять в дальнейших расчетах.

Если же получим небольшие значения коэффициентов тесноты связи, обычно ниже 0,7, и коэффициент существенности коэффициентов тесноты связи ниже, чем его табличное значение, то необходимо вернуться к первому этапу построения модели, т. е. к выбору факторов, найти неучтенные в модели, но существенно влияющие на результат факторы, ввести их в уравнение и повторить расчеты.

3. Кроме коэффициентов тесноты связи рассчитывают и коэффициент детерминации (D):

$$D = r^2 \cdot 100; \quad D = R^2(\eta)^2 100. \quad (2.10)$$

Коэффициент детерминации показывает, на сколько процентов учтенные в модели факторы объясняют вариацию (изменение) резуль- тативного показателя.

Например, $D = 0,948^2 \cdot 100 = 89,9$, т. е. на 89,9 % выбранные факто- ры объясняют вариацию результата. А на $100 - 89,9 = 10,1$ % – на из- менение резуль- тативного признака оказывают влияние неучтенные в модели факторы.

4. Скорректированный коэффициент детерминации равен

$$\bar{D} = 1 - (1 - R^2) \frac{n - 1}{n - m - 1}, \quad (2.11)$$

где n – число наблюдений;

m – число факторов уравнения, включая резуль- тативный.

Полученное выражение связывает два показателя тесноты связи – скорректированный и нескорректированный коэффициенты детер- минации. Скорректированный коэффициент детерминации применяется для решения двух задач:

- *оценки реальной тесноты связи между резуль- татом и фактора- ми;*

- *сравнения моделей с разным числом параметров.*

В первом случае обращают внимание на близость скорректирован- ного и нескорректированного коэффициентов детерминации. Если эти показатели велики и различаются незначительно, модель считается хорошей.

При сравнении разных моделей предпочтение при прочих равных условиях отдается той, у которой больше скорректированный коэффи- циент детерминации.

Следует отметить, что область применения скорректированного ко- эффициента детерминации ограничивается только этими задачами. Его

нельзя использовать в формулах, в которых применяется обычный коэффициент детерминации. Скорректированный коэффициент детерминации нельзя интерпретировать как долю вариации результата, объясненную вариацией факторов, включенных в модель регрессии.

Скорректированные коэффициенты детерминации близки к нескорректированным (исходным), что свидетельствует о хорошем качестве рассматриваемых моделей.

5. Для определения ценности модели в целом используют **критерий Фишера** (или **F-критерий**), который равен отношению общей дисперсии к остаточной:

$$F = \frac{\delta_{\text{общ}}^2}{\delta_{\text{ост}}^2}; F = \frac{R^2}{1-R^2} \frac{n-m-1}{m}. \quad (2.12)$$

Полученное значение критерия ($F_{\text{расч}}$) сравнивают с критическим (табличным) ($F_{\text{табл}}$) для принятого уровня значимости (α) и числа степеней свободы ($\nu_1 = m - 1$ и $\nu_2 = n - m$, где n – число наблюдений, m – число факторов уравнения, включая результативный). Если оно окажется больше соответствующего табличного значения (1,5), то данное уравнение статистически значимо, т. е. доля вариации, обусловленная регрессией, намного превышает случайную ошибку. Если же $F_{\text{расч}} < F_{\text{табл}}$, то модель неадекватно описывает реальный процесс и ее использовать для анализа и планирования нельзя, нужно пересмотреть выбор факторных показателей.

6. Для обобщающей оценки модели рассчитывают показатель **средней относительной ошибки аппроксимации** $\bar{\varepsilon}$, показывающий среднее отклонение расчетных значений зависимой переменной от соответствующих искомым величин:

$$\bar{\varepsilon} = \frac{1}{n} \sum \left| \frac{y_i - y_x}{y_i} \right| 100. \quad (2.13)$$

Функция имеет высокую точность, если $\bar{\varepsilon}$ не превышает 10 %. Если $\bar{\varepsilon}$ более 10 % и менее 20 % – модель имеет допустимую точность и может быть использована для анализа и прогноза.

Характеристики факторных показателей.

1. Кроме проверки ценности КМ в целом проверяют на существенность и каждый коэффициент регрессии, что делают с помощью **t-критерия Стьюдента**:

$$t_{a_j} = \frac{a_j}{\mu_{a_j}}, \quad (2.14)$$

где t_{a_j} – коэффициент существенности коэффициента регрессии;

μ_{a_j} – ошибка коэффициента регрессии.

Ошибка коэффициента регрессии определяется по формуле

$$\mu_{a_j} = \frac{\delta_{y_x y_i}}{\delta_{x_j} \sqrt{n}}, \quad (2.15)$$

где $\delta_{y_x y_i}$ – среднеквадратическое отклонение по данным уравнения.

Среднеквадратическое отклонение по данным уравнения определяется по формуле

$$\delta_{y_x y_i} = \sqrt{\frac{\sum (y_x - y_i)^2}{n}}. \quad (2.16)$$

Расчетные значения t -критерия Стьюдента сравнивают с критическими, которые определяют по таблице с учетом принятого уровня значимости ($\alpha = 0,10$; $\alpha = 0,05$ или $\alpha = 0,01$) и числа степеней свободы $\nu = n - m - 1$ (где n – число наблюдений; m – число факторов уравнения). Параметр признается значимым, если $t_{\text{расч}} \geq t_{\text{табл}}$ (иногда принимают 1,96).

Некоторые исследователи придерживаются мнения, что коэффициенты регрессии:

- 1) относительно значимы, если $1 < |t_{a_j}| < 2$;
- 2) значимы, если $2 < |t_{a_j}| < 3$;
- 3) сильно значимы, если $|t_{a_j}| > 3$.

Если есть коэффициент регрессии, для которого условие не выполняется, то из уравнения исключают тот фактор, коэффициент при котором незначим и имеет наименьшее значение t -критерия. После этого уравнение регрессии строится без исключенного фактора и снова про-

веряется значимость коэффициентов регрессии. Такой процесс длится до тех пор, пока все коэффициенты регрессии не окажутся значимыми, что свидетельствует о наличии в уравнении только существенных (действительно влияющих на результативный показатель) факторов.

2. Коэффициенты регрессии показывают эффективность каждого ресурса. Но так как факторные показатели КМ имеют различные единицы измерения, то коэффициенты регрессии несравнимы между собой. Вместе с тем часто требуется оценить роль факторных показателей в формировании результата и сравнить их между собой. Для этого используются **коэффициенты эластичности** (\mathcal{E}_{x_j}), которые определяются по формуле

$$\mathcal{E}_{x_j} = a_j \frac{\bar{x}_j}{\bar{y}}, \quad (2.17)$$

где a_j – коэффициент регрессии при j -м факторном признаке;

\bar{x}_j – среднее арифметическое значение факторного признака;

\bar{y} – среднее арифметическое значение результативного признака.

Коэффициент эластичности показывает, на сколько процентов изменится результативный показатель, если факторный показатель изменится на 1 %. Недостаток данного показателя заключается в том, что он применяется только при приблизительно одинаковой вариации факторов.

Коэффициенты эластичности имеют недостаток, который состоит в том, что они приемлемы, если вариации факторов одинаковы или почти одинаковы.

Допустим, $\bar{x}_1 = 20$ изменяется в диапазоне

$$(x_1^{\min} = 19) \leq x_1 \leq (x_1^{\max} = 22),$$

$\bar{x}_2 = 40$ изменяется в диапазоне

$$(x_2^{\min} = 25) \leq x_2 \leq (x_2^{\max} = 70).$$

Вариация первого факторного значения меньше, чем второго. Это значит, что первый фактор в меньшей степени объясняет изменение результативного показателя, чем второй.

3. Если вариация факторов отличается существенно, то для объяснения роли отдельных факторов в формировании результативного по-

казателя используются **β -коэффициенты** или **стандартизированные коэффициенты регрессии** β_{x_j} :

$$\beta_{x_j} = a_j \frac{\sigma_{x_j}}{\sigma_y}, \quad (2.18)$$

где a_j – коэффициент регрессии при j -м факторном признаке;

σ_{x_j} – стандартное (среднее квадратическое) отклонение j -го факторного признака;

σ_y – стандартное (среднее квадратическое) отклонение результативного признака.

Он показывает, на какую часть стандартного отклонения изменяется зависимая переменная с изменением фактора x_j на величину своего стандартного отклонения.

Если все β -коэффициенты суммировать и сумма будет больше единицы, то это означает, что значение результативного показателя увеличивается более быстрыми темпами, чем происходит прирост факторов. Если же данная сумма меньше единицы, то прирост результативного показателя отстает от темпов прироста факторов. Вышеизложенное касается и коэффициентов эластичности.

4. По рассчитанным β -коэффициентам и коэффициентам парной корреляции можно оценить индивидуальный вклад каждого факторного показателя в вариацию зависимой переменной. Для этой цели используются **показатели частной детерминации**:

$$d_{x_j} = \beta_{x_j} \cdot r_{yx_j}. \quad (2.19)$$

Сумма показателей частной детерминации равна коэффициенту детерминации $\left(\sum d_{x_j} = R^2 \right)$. Например, если построенная модель объясняет 89,9 % общей вариации переменной y , на долю фактора x_1 приходится 44,0 % ($d_{x_1} = 0,504 \cdot 0,873 = 0,440$); фактора x_2 – 37,3 % ($d_{x_2} = 0,432 \cdot 0,863 = 0,373$) и т. д.

Комбинируя факторы-ресурсы производства различным образом, можно обеспечить высокий уровень результатов. Причем в определенных пределах имеется возможность замещения одного ресурса другим.

2.4. Модели множественной линейной регрессии

Достаточно широкое распространение при оценке параметров моделей получил и метод максимального правдоподобия, базирующийся на критерии (принципе), согласно которому оптимальные оценки параметров обеспечивают максимум так называемой функции правдоподобия. В основе метода максимального правдоподобия (ММП) лежит исходное предположение о том, что «лучшим» оценкам $a_0^*, a_1^*, \dots, a_n^*$ «истинных» значений параметров эконометрической модели $\alpha_0, \alpha_1, \dots, \alpha_n$ должен соответствовать наиболее вероятный набор «фактических» значений ошибки $e_1^*, e_2^*, \dots, e_T^*$, рассматриваемых как «своего рода оценки» ее истинных значений $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$.

Таким образом, максимум произведения $p(e_1)p(e_2) \dots p(e_T)$ соответствует наиболее вероятному сочетанию значений $\varepsilon_t, t = \overline{1, T}$, обеспечиваемому «наилучшими» оценками параметров модели. При этом имеется в виду, что для произвольного набора значений фактической ошибки e_1, e_2, \dots, e_T произведение вероятностей $p(e_1)p(e_2) \dots p(e_T)$ в данном случае выражает вероятность совместного распределения их значений, соответствующих определенному набору оценок параметров a_0, a_1, \dots, a_n .

С учетом этого решение задачи оценки параметров ОЛСМ может быть получено в результате максимизации целевой функции следующего вида:

$$L(\varepsilon_1, \dots, \varepsilon_T) = p(\varepsilon_1) \cdot \dots \cdot p(\varepsilon_T) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} (y_t - \alpha_0 - \alpha_1 x_{1t} - \dots - \alpha_n x_{nt})^2 \right] \quad (2.20)$$

по неизвестным параметрам $\alpha_0, \alpha_1, \dots, \alpha_n$ и σ_ε^2 при заданных массивах исходных данных, выражаемых вектором известных значений зависимой переменной y_t и матрицей значений независимых факторов x размера $T(n+1)$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{pmatrix}; x = \begin{pmatrix} 1 & x_{11} & \dots & x_{n1} \\ 1 & x_{12} & \dots & x_{n2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1T} & \dots & x_{nT} \end{pmatrix}, \quad (2.21)$$

в которой столбец, состоящий из единиц, соответствует коэффициенту α_0 .

Целевая функция типа (2.4) называется **функцией максимального правдоподобия**.

Для получения оптимальных значений оценок параметров $a_0^*, a_1^*, \dots, a_n^*$ необходимо функцию (2.4) прологарифмировать, взяв производную по $\alpha_0, \alpha_1, \dots, \alpha_n$ и σ_e^2 .

После преобразований получим вектор оценок ММП коэффициентов ОЛСМ в следующем виде:

$$a^* = a = (x^T x)^{-1} x^T y, \quad (2.22)$$

а также оценку ММП дисперсии ошибки эконометрической модели:

$$\sigma_e^2 = \frac{1}{T} (y - x \cdot a)^T (y - x \cdot a) = \frac{1}{T} \sum_{i=1}^T e_i^2. \quad (2.23)$$

Заметим, что выражение (2.6) ничем не отличается от своего аналога, полученного с использованием МНК, а оценка дисперсии ошибки модели, полученная на основании выражения (2.7), является смещенной. Вследствие этого на практике используют несмещенную оценку дисперсии, определяемую следующим образом:

$$\sigma_e^2 = \frac{1}{T - n - 1} \sum_{i=1}^T e_i^2. \quad (2.24)$$

Следует также отметить, что при предположении о нормальном законе распределения ошибки эконометрической модели и ее свойствах оценки ее коэффициентов, полученные с использованием методов максимального правдоподобия и наименьших квадратов, совпадают.

Если же ошибки модели распределены по другому закону (например, Коши с тяжелыми хвостами, Стьюдента и т. п.), то, вообще говоря, выражения для оценки коэффициентов, полученные на основании

ММП, будут отличаться от их аналогов, полученных с использованием МНК.

3. ПОСТРОЕНИЕ И АНАЛИЗ ОБЩЕЙ ЛИНЕЙНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ В ПРЕДПОЛОЖЕНИИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ОШИБОК НАБЛЮДЕНИЯ

3.1. Свойства МНК-оценок.

3.2. Обобщенная регрессионная модель.

3.1. Свойства МНК-оценок

Теория статистического оценивания определяет качество оценок по свойствам несмещенности, эффективности и состоятельности и некоторым другим.

1. Оценка является *несмещенной*, если истинное значение параметра можно рассматривать как ее математическое ожидание:

$$M[a_i] = \alpha_i. \quad (3.1)$$

Если это не так, то оценка называется *смещенной*, а разность $M[a_i] - \alpha_i$ – смещением.

Несмещенность означает, что «в среднем» наша оценка соответствует параметру. Возможно, что полученное значение a_i будет удалено от α_i . При увеличении объема выборки n дисперсия x уменьшается, поскольку $\text{var}(x) = \sigma^2/n$, и соответственно становится меньше правдоподобие того, что a_i будет удалено от α_i .

Часто свойство несмещенности выполняется лишь с некоторой степенью «приблизительности» при достаточно больших объемах выборочных данных (при большом числе измерений T), в пределе при $T \rightarrow \infty$. В этом случае говорят, что оценки являются *асимптотически несмещенными*.

2. Если мы будем сравнивать две исследуемые оценки, то обратимся к их эффективности, основанной на соотношении их дисперсий. Та из оценок, которая имеет меньшую дисперсию, является более эффективной. Это свойство показано на рис. 3.1.

Оценка рассматривается как *эффективная*, если она характеризуется наименьшей дисперсией (дисперсия ошибки оценки минимальна)

среди всех других аналогичных оценок, полученных различными методами, способами.

$$\sigma_k^2(\alpha_i) = \min(\sigma_j^2(a_i)_j), \quad (3.2)$$

где $\sigma_j^2(a_i)$ – дисперсия оценки, полученной с использованием j -го метода оценивания.

Предположим, что имеются две оценки параметра α_0 , рассчитанные на основе одной и той же информации (рис. 3.1).

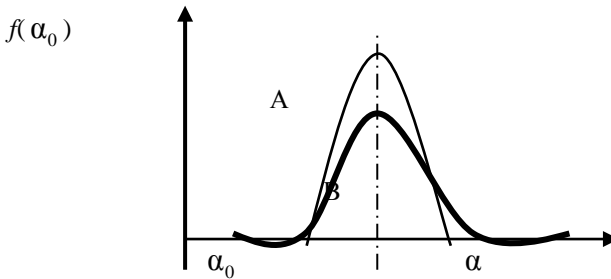


Рис. 3.1. Оценки параметра α_0

Обе оценки A и B являются несмещенными оценками α_0 , однако оценка A является более эффективной: она имеет дисперсию меньшую, чем оценка B .

В этом случае говорят, что оценки являются **асимптотически эффективными**.

3. Еще одним асимптотическим свойством является **состоятельность**. Состоятельной называется такая оценка, которая дает точное значение для большой выборки независимо от входящих в нее конкретных наблюдений. Это свойство записывается следующим образом:

$$p \lim(a_i^{(T)}) = \alpha_i, \quad (3.3)$$

где p – вероятность события, заключенного в фигурные скобки.

На рис. 3.2 показано, как при различном объеме выборки может выглядеть распределение вероятностей (состоятельная оценка, смещенная на малой выборке).

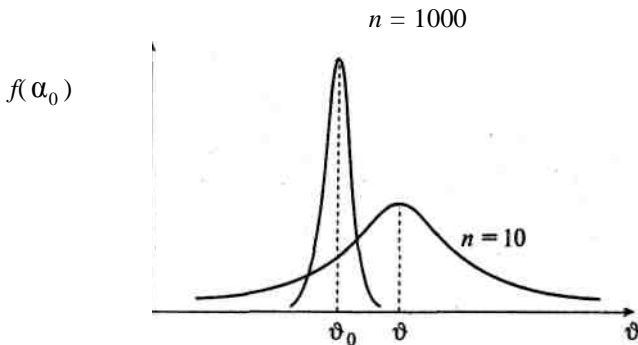


Рис. 3.2. Распределение вероятностей при различном объеме выборки

Однако следует проявлять определенную осторожность. Дело в том, что при конечных объемах выборки интуитивно предполагается: чем больше значение T , тем ближе полученная оценка параметра к его истинному значению и тем меньше ее ошибка. Но увеличение объема выборки ведет к уменьшению величины смещения только в том случае, если закономерности рассматриваемых процессов на «старом» и «добавленном» временных интервалах полностью идентичны (однородная выборка) и, таким образом, построенные на соответствующей этим интервалам информации модели будут малоразличимы.

В эконометрике часто приходится сталкиваться с такой ситуацией, когда на отдельных временных участках закономерности процессов различаются. Это может быть вызвано, например, начавшимся воздействием на зависимую переменную y нового фактора, изменением характера взаимосвязей между рассматриваемыми переменными в связи с изменением их масштабов (действие диалектического закона перехода «количества» в «качество») и по другим причинам. В этом случае увеличение числа измерений не ведет к автоматическому «росту качества» оценок параметров модели.

Состоятельность является «более удобным» при анализе свойств, чем асимптотическая несмещенность, поскольку оно часто автоматически сохраняется при преобразованиях рассматриваемых переменных и некоторых операциях с ними.

3.2. Обобщенная регрессионная модель

Для получения несмещенных, эффективных и состоятельных оценок параметров модели необходимо выполнение следующих предпосылок:

1. Остатки e_i ($i = 1, 2, \dots, n$) есть величина случайная, а факторы x_1, x_2, \dots, x_p – величины неслучайные. Это означает, что вектор e – случайный вектор, а матрица значений факторов x – неслучайная (детерминированная).

2. Математическое ожидание случайного остатка равно нулю:

$$M(E_i) = 0 \quad (i = 1, 2, \dots, n). \quad (3.4)$$

3. Дисперсия случайного остатка одинакова для всех наблюдений результата y :

$$\sigma^2(E_i) = \sigma^2(E_j) = \sigma^2 = \text{const} \quad (i = 1, 2, \dots, n). \quad (3.5)$$

Это условие называется условием *гомоскедастичности*. Напротив, если эта дисперсия не постоянна, то такое явление называют *гетероскедастичностью остатков*.

Выполнение этой предпосылки может проверяться разными методами. Ниже рассмотрена процедура проверки предпосылки методом Голдфелда – Квандта.

4. Возмущения не коррелированы (не зависят друг от друга) между собой. Это означает, что ковариация между отдельными возмущениями e_i и e_j ($i \neq j$) равна нулю:

$$r_{E_i E_j} = 0. \quad (3.6)$$

Проверка выполнения предпосылки 4 с помощью d -статистики Дарбина – Уотсона рассмотрена ниже.

5. Случайные остатки e_i есть нормально распределенная случайная величина, а вектор E – нормально распределенный случайный вектор:

$$\varepsilon \sim N_n(0; \sigma_\varepsilon^2 I_n). \quad (3.7)$$

Обоснованием такого допущения служит центральная предельная теорема теории вероятностей, согласно которой сумма большого числа случайных величин имеет приближенно нормальное распределение независимо от индивидуального распределения слагаемых. Отклоне-

ние фактических значений результата y от теоретических вызывается, как правило, множеством случайных и неучтенных факторов, каждый из которых не оказывает доминирующего влияния. Поэтому нормальное распределение является приемлемой моделью суммарной погрешности, т. е. возмущения.

Первые четыре условия известны как условия Гаусса – Маркова.

Теорема Гаусса – Маркова. *Если регрессионная модель удовлетворяет предпосылкам 1–4, то оценки a_0, a_1, \dots, a_p имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.*

Таким образом, оценки a_0, a_1, \dots, a_p в определенном смысле являются наиболее эффективными линейными оценками параметров $\alpha_0, \alpha_1, \dots, \alpha_n$.

При моделировании реальных экономических процессов нередко сталкиваемся с ситуациями, в которых условия классической линейной модели регрессии оказываются нарушенными. В частности, могут не выполняться предпосылки 3 и 4 о том, что случайные остатки модели имеют постоянную дисперсию и не коррелированы между собой.

В отличие от классической, в обобщенной модели ковариации и дисперсии объясняющие переменные могут быть произвольными. В этом состоит суть обобщения регрессионной модели.

4. МЕТОДЫ ПОСТРОЕНИЯ ОБЩЕЙ ЛИНЕЙНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ ПРИ НАРУШЕНИИ ТРАДИЦИОННЫХ ПРЕДПОЛОЖЕНИЙ ОТНОСИТЕЛЬНО ОШИБОК НАБЛЮДЕНИЙ

4.1. Тесты гетероскедастичности: графический анализ остатков, тест Голдфелда – Квандта, тест Уайта.

4.2. Экономические модели с автокоррелированными ошибками. Анализ автокорреляции ошибок на основе статистики и теста Дарбина – Уотсона. Процедура Кохрейна – Оркатта.

4.1. Тесты гетероскедастичности: графический анализ остатков, тест Голдфелда – Квандта, тест Уайта

На практике гетероскедастичность не так уж и редка. На рис. 4.1 приведены два примера линейной регрессии – зависимости потребления C от дохода I :

$$C = a_0 + a_1 I + e.$$

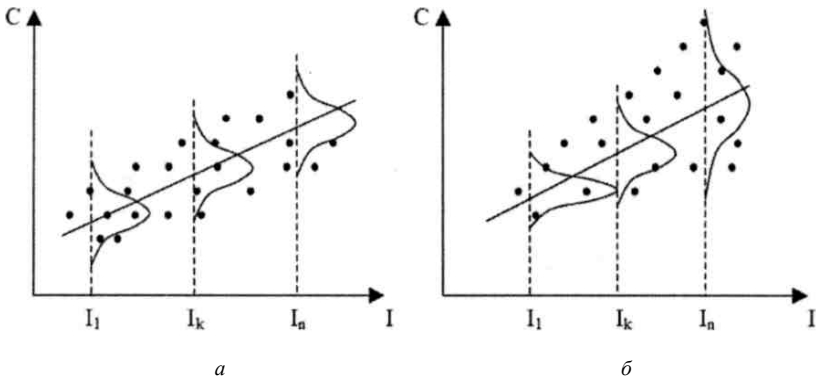


Рис. 4.1. Зависимость потребления C от дохода I

В обоих случаях с ростом дохода растет среднее значение потребления. Но если на рис. 4.1, *а*, дисперсия потребления остается одной и той же для различных уровней дохода, то на рис. 4.1, *б*, при аналогичной зависимости среднего потребления от дохода дисперсия потребления не остается постоянной, а увеличивается с ростом дохода. Фактически это означает, что во втором случае субъекты с большим доходом в среднем потребляют больше, чем субъекты с меньшим доходом, и, кроме того, разброс в их потреблении более существенен для большего уровня дохода. Фактически люди с большими доходами имеют больший простор для распределения своего дохода. Реальность данной ситуации не вызывает сомнений. Разброс значений потребления вызывает разброс точек наблюдения относительно линии регрессии, что и определяет дисперсию случайных отклонений.

Виды гетероскедастичности.

1. *Истинная гетероскедастичность* вызывается непостоянством дисперсии случайной компоненты, ее зависимостью от различных факторов. Наиболее распространенный случай истинной гетероскедастичности: дисперсия растет с ростом одного из факторов.

2. *Ложная гетероскедастичность* вызывается ошибочной спецификацией модели регрессии.

Последствия гетероскедастичности.

При гетероскедастичности последствия применения МНК будут следующими.

1. Оценки коэффициентов по-прежнему остаются несмещенными и линейными.

2. Оценки не будут эффективными (т. е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра). Они не будут даже асимптотически эффективными. Увеличение дисперсии оценок снижает вероятность получения максимально точных оценок.

3. Дисперсии оценок будут рассчитываться со смещением. Смещенность появляется вследствие того, что необъясненная уравнением

регрессии дисперсия $S^2 = \frac{\sum e_i^2}{m - n - 1}$ (m – число объясняющих пере-

менных), которая используется при вычислении оценок дисперсий всех коэффициентов, не является более несмещенной.

4. Вследствие вышесказанного все выводы, получаемые на основе соответствующих t - и F -статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получаемые при стандартных проверках качества оценок могут быть ошибочными и приводить к неверным заключениям по построенной модели. Вполне вероятно, что стандартные ошибки коэффициентов будут занижены, а следовательно, t -статистики будут завышены. Это может привести к признанию статистически значимыми коэффициентов, таковыми на самом деле не являющимися.

Обнаружение гетероскедастичности.

В ряде случаев на базе знаний характера данных появление проблемы гетероскедастичности можно предвидеть и попытаться устранить этот недостаток еще на этапе спецификации. Однако значительно чаще эту проблему приходится решать после построения уравнения регрессии.

Естественно, не существует какого-либо однозначного метода определения гетероскедастичности. 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 Уайта.

Графический анализ остатков.

Использование графического представления отклонений позволяет определить, с наличием гетероскедастичности. В этом случае по оси абсцисс откладывается объясняющая переменная x либо линейная комбинация объясняющих переменных $y = a_0 + a_1x_1 + \dots + a_mx_m$, а по оси ординат либо отклонения e_i , либо их квадраты.

Примеры таких графиков приведены на рис. 4.2.

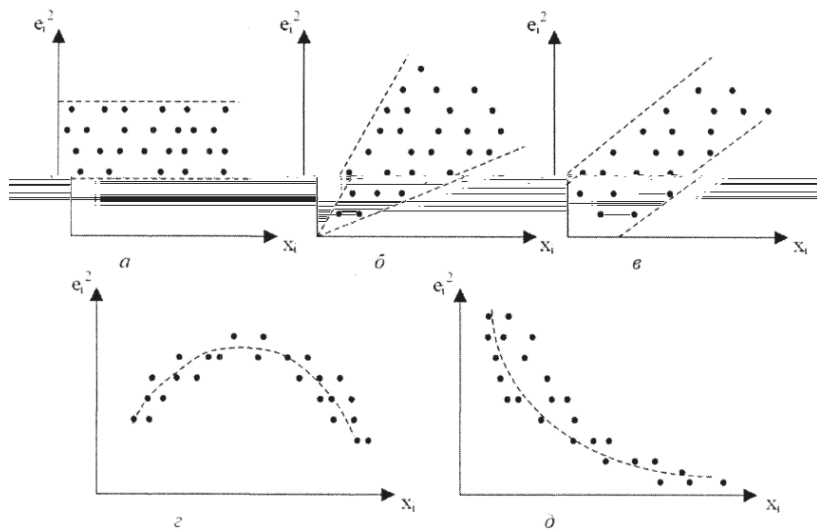


Рис. 4.2. Точечный график зависимости квадрата остатков e^2 от объясняющей переменной x

На рис. 4.2, *a*, все отклонения e^2_i находятся внутри полуполосы постоянной ширины, параллельной оси абсцисс, т. е. в этом случае мы находимся в условиях гомоскедастичности.

Ситуации, представленные на рис. 4.2, *б* – *д*, отражают большую вероятность наличия гетероскедастичности для рассматриваемых статистических данных.

Отметим, что графический анализ отклонений является удобным и достаточно надежным в случае парной регрессии. При множественной регрессии графический анализ возможен для каждой из объясняющих переменных x_i отдельно. Такой анализ наиболее целесообразен при большом количестве объясняющих переменных.

Тест Голдфелда – Квандта.

Тест Голдфелда – Квандта состоит в следующем:

1. Все n остатков упорядочиваются по возрастанию значений фактора x_j .
2. В упорядоченном ряду выбирают k первых и k последних остатков, при этом k должно быть больше числа факторов, включенных в

модель. Обычно принимают $k \approx n/3$. Центральные остатки, таким образом, исключаются из рассмотрения.

3. По каждой из групп выбранных остатков определяется сумма их квадратов:

$$\begin{aligned} SS_1 &= e_1^2 + e_2^2 + \dots + e_k^2 \text{ и} \\ SS_2 &= e_{n-k+1}^2 + e_{n-k+2}^2 + \dots + e_n^2. \end{aligned} \quad (4.1)$$

4. Рассчитывается F -статистика Фишера по формуле $F = SS_1 / SS_3$, если $SS_1 > SS_3$, или по формуле $F = SS_3 / SS_1$, если $SS_3 > SS_1$.

5. Статистическая гипотеза об одинаковой дисперсии возмущений не отклоняется, если F -статистика не превышает табличное значение F -критерия Фишера для принятого уровня значимости α и чисел степеней свободы числителя и знаменателя $\nu_1 = \nu_2 = k - m - 1$.

6. Если $F_{\text{набл}} > F_{\text{кр}} = F_{\alpha, H_1, H_2}$, то гипотеза об отсутствии гетероскедастичности отклоняется (здесь α – выбранный уровень значимости).

Естественным является вопрос, какими должны быть размеры подвыборок для принятия обоснованных решений. Для парной регрессии Голдфелд и Квандт предлагают следующие пропорции: $n = 32, k = 11$; $n = 60; k = 22$.

Тест Уайта.

Тест Уайта состоит в следующем:

1. Строится уравнение регрессии: $\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_{i3}$ и вычисляются остатки $e_i = y_i - \hat{y}_i, i = \overline{1, n}$.

2. Строится квадратичная функция, включающая все факторы, а также их попарные произведения:

$$\begin{aligned} e_i^2 &= \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i1}^2 + \alpha_5 x_{i2}^2 + \\ &+ \alpha_6 x_{i3}^2 + \alpha_7 x_{i1} x_{i2} + \alpha_8 x_{i1} x_{i3} + \alpha_9 x_{i2} x_{i3} + \eta_i \end{aligned} \quad (4.2)$$

При проверке по тесту Уайта: гетероскедастичность случайных остатков имеется, если вся функция (4.2) значима по F -критерию Фишера. Если фактическое значение критерия больше табличного, то гипотеза о гомоскедастичности остатков отклоняется.

Обобщенный (взвешенный) МНК.

При наличии гетероскедастичности используют обобщенный (взвешенный) МНК. Суть метода состоит в уменьшении вклада наблюдений, имеющих большую дисперсию, в результаты расчета. Пусть в исходной модели $y = \alpha_0 + \alpha_1 x + \varepsilon$ случайные компоненты гетероскедастичны, т. е. $M(\varepsilon) = 0, D(\varepsilon) = \sigma^2$. Допустим, что σ^2 в каждом наблюдении известны. Разделим каждое наблюдение на соответствующее ему значение σ и получим преобразованную модель вида:

$$y/\sigma = \alpha_0/\sigma + \alpha_1 x/\sigma + \varepsilon/\sigma.$$

Для полученной модели будут выполнены условия гомоскедастичности, так как $M(\varepsilon/\sigma) = 0, D(\varepsilon/\sigma) = 1$. Оценки параметров преобразованной модели находят обычным МНК. На практике σ^2 неизвестны, поэтому их заменяют оценками e , т. е. оценивается обычным МНК преобразованная модель: $y/e = \alpha_0/e + \alpha_1 x/e + \varepsilon/e$.

Для экономических данных σ часто пропорциональна значениям объясняющей переменной x , поэтому оценивается обычным МНК преобразованная модель: $y/x = \alpha_0/x + \alpha_1 + \varepsilon/x$. Коэффициент при $1/x$ будет эффективной оценкой α_0 , а постоянный член – эффективной оценкой α_1 . При применении взвешенного МНК оценки параметров будут несмещенными, кроме того, иметь меньшую дисперсию, чем невзвешенные оценки.

4.2. Экономические модели с автокоррелированными ошибками.

Анализ автокорреляции ошибок на основе статистики и теста Дарбина – Уотсона. Процедура Кохрейна – Оркатта

Автокорреляция возмущений бывает положительной или отрицательной.

Положительная автокорреляция проявляется в том, что завышенные значения возмущений предыдущих наблюдений результата y приводят к завышению возмущений последующих наблюдений. На графике временного ряда остатков регрессии это выражается, например, в чередовании зон положительных и отрицательных остатков (рис. 4.3). Графически положительная автокорреляция выражается в чередовании зон, где наблюдаемые значения оказываются выше объясненных (предсказанных), и зон, где наблюдаемые значения ниже.

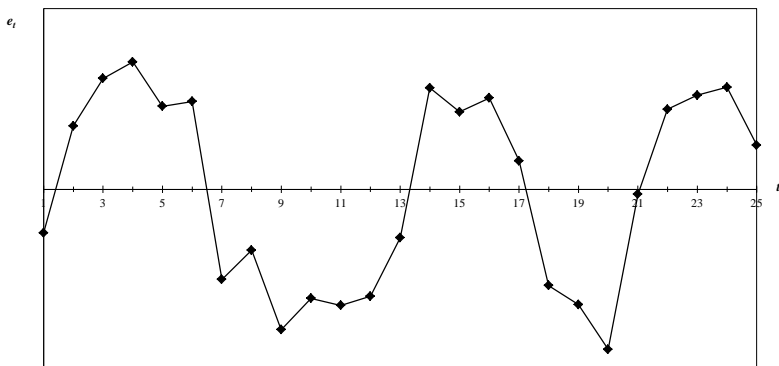


Рис. 4.3. Модель регрессии с положительной автокорреляцией возмущений

При **отрицательной автокорреляции**, наоборот, завышенные значения возмущений предыдущих наблюдений занижают возмущения последующих наблюдений, а остатки регрессии «слишком часто» меняют знак (рис. 4.4). Графически это выражается в том, что результаты наблюдений y_t «слишком часто» «перескакивают» через график объясненной части t .

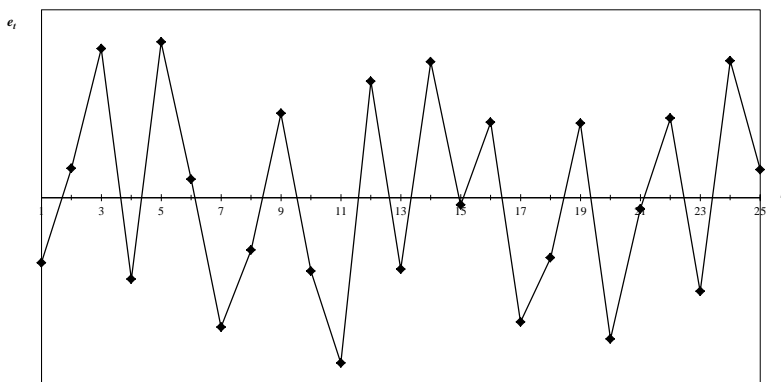


Рис. 4.4. Модель регрессии с отрицательной автокорреляцией возмущений

Среди основных причин, вызывающих появление автокорреляции, можно выделить ошибки спецификации, инерцию в изменении экономических показателей, эффект паутины, сглаживание данных.

1. *Ошибки спецификации.* Неучет в модели какой-либо важной объясняющей переменной либо неправильный выбор формы зависимости обычно приводит к системным отклонениям точек наблюдений от линии регрессии, что может привести к автокорреляции.

Проиллюстрируем это следующим примером. Анализируется зависимость предельных издержек MC от объема выпуска Q . Если для ее описания вместо реальной квадратичной модели $MC = \beta_0 + \beta_1 Q + \beta_2 Q^2 + \varepsilon$ выбрать линейную модель $MC = \beta_0 + \beta_1 Q + \varepsilon$, то совершается ошибка спецификации. Ее можно рассматривать как неправильный выбор формы модели или как отбрасывание значимой переменной при линеаризации указанных моделей. Последствия данной ошибки выразятся в системном отклонении точек наблюдений от прямой регрессии (рис. 4.5) и существенном преобладании последовательных отклонений одинакового знака над соседними отклонениями противоположных знаков. Налицо типичная картина, характерная для положительной автокорреляции.

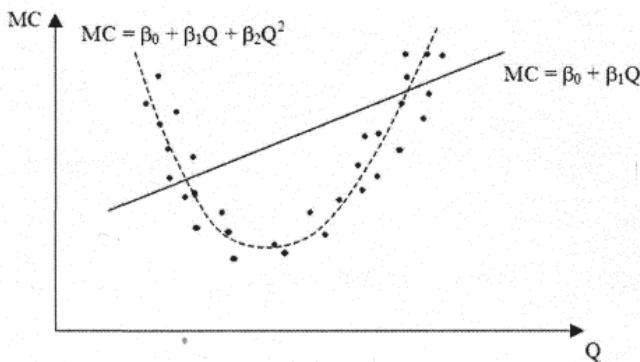


Рис. 4.5. Зависимость издержек от объема выпуска

2. *Инерция.* Многие экономические показатели (например, инфляция, безработица, ВВП и т. п.) обладают определенной цикличностью, связанной с волнообразностью деловой активности. Действительно, экономический подъем приводит к росту занятости, сокращению инфляции, увеличению ВВП и т. д. Этот рост продолжается до тех пор, пока изменение конъюнктуры рынка и ряда экономических характеристик не приведет к замедлению роста, затем остановке и движению

назад рассматриваемых показателей. В любом случае эта трансформация происходит не мгновенно, а обладает определенной инертностью.

3. *Эффект паутины.* Во многих сферах экономики экономические показатели реагируют на изменение экономических условий с запаздыванием (временным лагом). Например, предложение сельскохозяйственной продукции реагирует на изменение цены с запаздыванием (равным периоду созревания урожая). Большая цена сельскохозяйственной продукции в прошлом году вызовет (скорее всего) ее перепроизводство в текущем году, а следовательно, цена на нее снизится и т. д. В этой ситуации нельзя предполагать случайность отклонений друг от друга.

4. *Сглаживание данных.* Нередко данные по некоторому продолжительному временному периоду получают усреднением данных по составляющим его подынтервалам. Это может привести к определенному сглаживанию колебаний, которые имелись внутри рассматриваемого периода, что, в свою очередь, может послужить причиной автокорреляции.

Последствия автокорреляции.

Последствия автокорреляции в определенной степени сходны с последствиями гетероскедастичности. Среди них при применении МНК обычно выделяются следующие.

1. Оценки параметров, оставаясь линейными и несмещенными, перестают быть эффективными. Следовательно, они перестают обладать свойствами наилучших линейных несмещенных оценок.

2. Дисперсии оценок являются смещенными. Нередко дисперсии, вычисляемые по стандартным формулам, являются заниженными, что приводит к увеличению t -статистик. Это может привести к признанию статистически значимыми объясняющие переменные, которые в действительности таковыми могут и не являться.

3. Оценка дисперсии регрессии
$$s^2 = \frac{\sum e_t^2}{m - n - 1}$$
 является смещенной

оценкой истинного значения σ^2 , во многих случаях занижая его.

4. В силу вышесказанного выводы по t - и F -статистикам, определяющим значимость коэффициентов регрессии и коэффициента детерминации, возможно, будут неверными. Вследствие этого ухудшаются прогнозные качества модели.

Обнаружение автокорреляции.

Рассмотрим возможные методы определения автокорреляции.

Графический метод.

Существует несколько вариантов графического определения автокорреляции. Один из них, увязывающий отклонения e_t с моментами t их получения (их порядковыми номерами i), приведен на рис. 4.6. Это так называемые последовательно-временные графики. В этом случае по оси абсцисс обычно откладываются либо момент получения статистических данных, либо порядковый номер наблюдения, а по оси ординат – отклонения ε_t (либо оценки отклонений e_t).

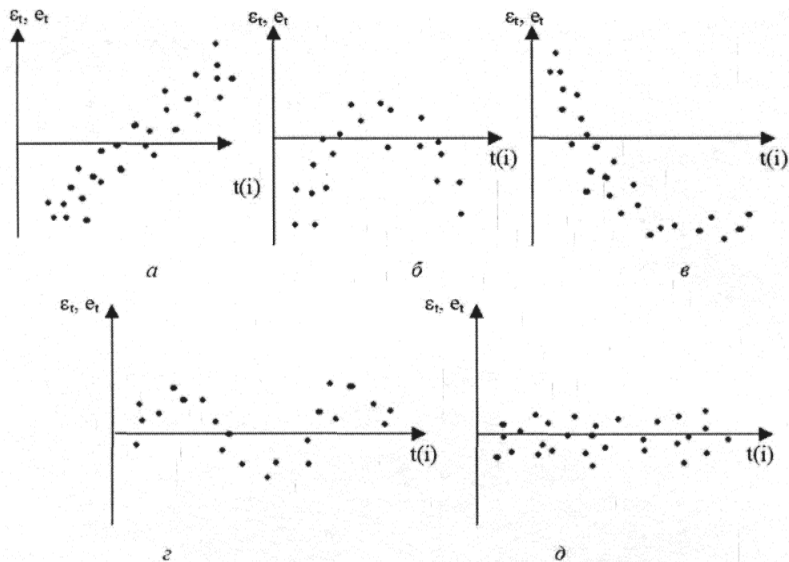


Рис. 4.6. Точечный график зависимости отклонения ε_t от времени t

Естественно предположить, что на рис. 4.6, *а – г*, имеются определенные связи между отклонениями, т. е. автокорреляция имеет место. Отсутствие зависимости на рис. 4.6, *д*, скорее всего, свидетельствует об отсутствии автокорреляции.

Критерий Дарбина – Уотсона.

Наиболее известным критерием обнаружения автокорреляции первого порядка является критерий Дарбина – Уотсона. Статистика DW Дарбина – Уотсона приводится во всех эконометрических пакетах как важнейшая характеристика качества регрессионной модели. Тест Дарбина – Уотсона основан на простой идее: если корреляция ошибок ре-

грессии не равна нулю, то она присутствует и в остатках регрессии e_t , получающихся в результате применения обычного метода наименьших квадратов. В тесте Дарбина – Уотсона для оценки корреляции используется статистика вида

$$DW = d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} \approx 2(1 - r_{e_t, e_{t-1}}). \quad (4.3)$$

Таким образом, $0 < DW < 4$ и его значения могут указать на наличие либо отсутствие автокорреляции. Действительно, если $r_{e_t, e_{t-1}} \approx 0$ (автокорреляция отсутствует), то $DW \approx 2$. Если $r_{e_t, e_{t-1}} \approx 1$ (положительная автокорреляция), то $DW \approx 0$. Если $r_{e_t, e_{t-1}} \approx -1$ (отрицательная автокорреляция), то $DW \approx 4$.

Для более точного определения, какое значение DW свидетельствует об отсутствии автокорреляции, а какое о ее наличии, была построена таблица критических точек распределения Дарбина – Уотсона. По ней для заданного уровня значимости α , числа наблюдений n и количества объясняющих переменных m определяются два значения: d_H – нижняя граница и d_B – верхняя граница.

Общая схема критерия Дарбина – Уотсона будет следующей:

1. По построенному эмпирическому уравнению регрессии $y_t = b_0 + b_1 x_{t1} + \dots + b_m x_{tm}$ определяются значения отклонений $e_t = y_t - \hat{y}_t$ каждого наблюдения $t, t = \overline{1, T}$.

2. По формуле (4.2) рассчитывается статистика DW .

3. По таблице критических точек Дарбина – Уотсона определяют два числа d_H и d_B и осуществляют выводы по следующей схеме:

$0 \leq DW < d_H$ – существует положительная автокорреляция;

$d_H \leq DW < d_B$ – вывод о наличии автокорреляции не определен;

$d_B \leq DW < 4 - d_B$ – автокорреляция отсутствует;

$4 - d_B \leq DW < 4 - d_H$ – вывод о наличии автокорреляции не определен;

$4 - d_H \leq DW < 4$ – существует отрицательная автокорреляция.

Изобразим результат Дарбина – Уотсона графически (рис. 4.7).



Рис. 4.7. Графическая интерпретация результатов критерия Дарбина – Уотсона

Для d -статистики находят верхнюю d_B и нижнюю d_H границы на уровнях значимости $\alpha = 0,01; 0,025$ и $0,05$ (норматив). Недостатками критерия Дарбина – Уотсона является наличие области неопределенности критерия, а также то, что критические значения d -статистики определены для объемов выборки не менее 15. Тем не менее тест Дарбина – Уотсона является наиболее употребляемым.

Методы устранения автокорреляции.

Процедура, которую следует принять для устранения автокорреляции, будет зависеть от характера зависимости между значениями случайного члена. Наибольшее внимание уделяется так называемой авторегрессионной схеме первого порядка, так как она интуитивно правдоподобна, но для того, чтобы было целесообразным ее использование в более сложных моделях, оснований обычно не хватает.

Предположим, что истинная модель задается выражением

$$y_t = \alpha + \beta x_t + u_t, \quad (4.4)$$

так что наблюдение $t - 1$ формируется как

$$y_{t-1} = \alpha + \beta x_{t-1} + u_{t-1}. \quad (4.5)$$

Теперь вычтем из обеих частей уравнения (4.4) умноженное на ρ соотношение (4.5) и получим:

$$y_{t-1} - \rho y_{t-1} = \alpha(1 - \rho) + \beta(x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1}). \quad (4.6)$$

Обозначим

$$\tilde{y}_t = y_{t-1} - \rho y_{t-1}, \tilde{x}_t = x_t - \rho x_{t-1}, \varepsilon_t = u_t - \rho u_{t-1}, \tilde{q} = 1 - \rho.$$

Тогда формулу (4.6) можно переписать так:

$$\tilde{y}_t = \alpha \tilde{q} + \beta \tilde{x}_t + \varepsilon_t. \quad (4.7)$$

Мы предположили, что ρ известно. Тогда можно вычислить величины $\tilde{y}_t, \tilde{q}_t, \tilde{x}_t$ для наблюдений, включающих от 2 до T исходных данных. Если теперь оценить регрессию между $\tilde{y}_t, \tilde{q}_t, \tilde{x}_t$, то будут получены оценки α и β , не связанные с проблемой автокорреляции, поскольку, согласно предположению, значения ε не зависят друг от друга.

Остается, однако, небольшая проблема. Если в выборке нет данных, предшествующих первому наблюдению, то мы не сможем вычислить \tilde{y}_1, \tilde{x}_1 и потеряем первое наблюдение. Число степеней свободы уменьшается на единицу, и это вызовет потерю эффективности, которая может в небольших выборках перевесить повышение эффективности от устранения автокорреляции.

Эту проблему, к счастью, можно довольно легко обойти, пользуясь так называемой поправкой Прайса – Уинстена (Prais, Winsten, 1954). Случайный член ε , согласно определению, не зависит от значения и в любом предшествующем наблюдении. В частности, все величины $\varepsilon_2, \dots, \varepsilon_T$ не зависят от u_1 . Следовательно, если при устранении автокорреляции все другие наблюдения преобразуются, то не требуется преобразовывать первое наблюдение. Можно сохранить его, включив в новую схему, полагая, что $\tilde{y}_1 = y_1, \tilde{x}_1 = x_1, \tilde{q}_1 = 1$.

Мы можем таким способом спасти первое наблюдение, но здесь есть небольшая проблема, которую требуется решить. Если ρ велико, то первое наблюдение будет оказывать непропорционально большое воздействие на оценки, исчисленные по уравнению регрессии. Чтобы нейтрализовать этот эффект, уменьшим вес данного наблюдения умножением его на величину $\sqrt{1-\rho^2}$, полагая, что $\tilde{y}_1 = \sqrt{1-\rho^2} y_1$, $\tilde{x}_1 = \sqrt{1-\rho^2} x_1$ и $\tilde{q}_1 = \sqrt{1-\rho^2}$.

Параметр ρ , как правило, неизвестен. Его оценку производят, исследуя взаимосвязь остатков, полученных после применения к модели регрессии обычного МНК.

Чтобы оценить значение ρ , а затем подобрать другие параметры модели, используют оценку $\hat{\rho}$. Для оценки ρ чаще всего используется процедура, предложенная Кохрейном (Cochrane) и Оркаттом (Orcutt).

Метод Кохрейна – Оркатта является итерационным. Каждая итерация дает лучшую оценку ρ , чем предыдущая.

1. На первом этапе с помощью метода наименьших квадратов оцениваются параметры простой линейной регрессии: $y_t = \alpha + \beta x_t + \varepsilon_t$.

2. Вычисляем остатки, полученные из этого уравнения:

$$e_t = y_t - \hat{y}_t = y_t - \hat{\alpha} - \hat{\beta}x_t.$$

3. Оценивается регрессионная зависимость e_t от e_{t-1} и коэффициент при e_{t-1} представляет собой оценку ρ , т. е. используется для оценки параметров модели:

$$e_t = \rho e_{t-1} + v_t. \quad (4.8)$$

Оценка, полученная методом наименьших квадратов, является начальной оценкой значения ρ :

$$\rho = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}. \quad (4.9)$$

4. С этой оценкой ρ уравнение (4.4) преобразуется в (4.9), оценивание которого позволяет получить пересмотренные оценки α , β , т. е. эта оценка дальше используется для нахождения обобщенных разностей: $y'_t = y_t - \hat{\rho}y_{t-1}$, $x'_t = x_t - \hat{\rho}x_{t-1}$. Для преобразованной модели $y'_t = \alpha(1 - \hat{\rho}) + \beta x'_t + \varepsilon_t$ проводится регрессионный анализ, дающий уточненные оценки $\hat{\alpha}$ и $\hat{\beta}$ наклона и свободного члена. На этом первая итерация заканчивается.

5. Уточненные оценки из первой итерации подставляем в начальное уравнение $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$, из которого находим новые значения остатков e_t и возвращаемся к этапу 3.

Новые остатки подставляем в уравнение (4.9), вычисляем новую оценку ρ и продолжаем итерации дальше. Этот процесс заканчивается, когда очередной шаг не дает существенного изменения величины $\hat{\rho}$.

После завершения итераций для окончательно преобразованной модели (см. уравнение (4.9)) можно провести стандартный регрессионный анализ и сделать выводы о связи между величинами y и x . Найденную модель также можно использовать для прогноза будущих значений y по известным величинам x .

5. ПОСТРОЕНИЕ И АНАЛИЗ ОБЩЕЙ ЛИНЕЙНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ В УСЛОВИЯХ МУЛЬТИКОЛЛИНЕАРНОСТИ ФАКТОРОВ

5.1. Мультиколлинеарность факторов: причины и эффекты.

5.2. Методы построения ОЛСМ в условиях мультиколлинеарности факторов.

5.1. Мультиколлинеарность факторов: причины и эффекты

Еще одной серьезной проблемой при построении моделей множественной линейной регрессии по МНК является **мультиколлинеарность** – линейная взаимосвязь двух или нескольких объясняющих переменных. Если объясняющие переменные связаны строгой функциональной зависимостью, то говорят о *совершенной мультиколлинеарности*. На практике можно столкнуться с очень высокой (или близкой к ней) мультиколлинеарностью – сильной корреляционной зависимостью между объясняющими переменными.

Мультиколлинеарность может быть проблемой лишь в случае множественной регрессии. В случае мультиколлинеарности выводы по коэффициентам и по самому уравнению регрессии будут ненадежными.

Совершенная мультиколлинеарность является скорее теоретическим примером. Реальна же ситуация, когда между объясняющими переменными существует довольно сильная корреляционная зависимость, а не строгая функциональная. Такая зависимость называется *несовершенной мультиколлинеарностью*. Она характеризуется высоким коэффициентом корреляции между соответствующими объясняющими переменными. Причем, если его значение по абсолютной величине близко к единице, то говорят о почти совершенной мультиколлинеарности. В любом случае мультиколлинеарность затрудняет разделение влияния объясняющих факторов на поведение зависимой переменной и делает оценки коэффициентов регрессии ненадежными. Данный вывод наглядно подтверждается с помощью диаграммы Венна (рис. 5.1).

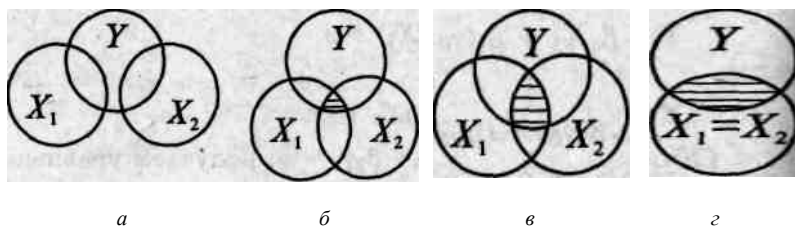


Рис. 5.1. Диаграмма Венна

На рис. 5.1, а, коррелированность между объясняющими переменными x_1 и x_2 отсутствует и влияние каждой из них на y находит отражение в наложении кругов x_1 и x_2 на круг y . По мере усиления линейной зависимости между x_1 и x_2 соответствующие круги все больше накладываются друг на друга. Заштрихованная область отражает совпадающие части влияния x_1 и x_2 на y . На рис. 1, з, при совершенной мультиколлинеарности невозможно разграничить степени индивидуального влияния объясняющих переменных x_1 и x_2 на зависимую переменную y .

Количественные меры мультиколлинеарности. Как известно, при выполнении определенных предпосылок МНК дает наилучшие линейные несмещенные оценки. Причем свойство несмещенности и эффективности оценок остается в силе, даже если несколько коэффициентов регрессии оказываются статистически незначимыми. Однако несмещенность фактически означает лишь то, что при многократном повторении наблюдений (при постоянных объемах выборок) за исследуемыми величинами средние значения оценок стремятся к их истинным значениям. К сожалению, повторять наблюдения в одинаковых условиях в экономике практически невозможно. Поэтому это свойство ничего не гарантирует в каждом конкретном случае. Наименьшая возможная дисперсия совсем не означает, что дисперсия оценок будет мала по сравнению с самими оценками. В ряде случаев такая дисперсия достаточно велика, чтобы оценки коэффициентов стали статистически незначимыми.

Обычно выделяются следующие **последствия мультиколлинеарности**:

1. Большие дисперсии (стандартные ошибки) оценок. Это затрудняет нахождение истинных значений определяемых величин и расширяет интервальные оценки, ухудшая их точность.

2. Уменьшаются t -статистики коэффициентов, что может привести

к неоправданному выводу о существенности влияния соответствующей объясняющей переменной на зависимую переменную.

3. Оценки коэффициентов по МНК и их стандартные ошибки становятся очень чувствительными к малейшим изменениям данных, т. е. они становятся неустойчивыми.

4. Затрудняется определение вклада каждой из объясняющих переменных в объясняемую уравнением регрессии дисперсию зависимой переменной.

5. Возможно получение неверного знака у коэффициента регрессии.

Существует несколько **признаков, по которым может быть установлено наличие мультиколлинеарности:**

1. Коэффициент детерминации R^2 достаточно высок, но некоторые из коэффициентов регрессии статистически незначимы, т. е. они имеют низкие t -статистики.

2. Парная корреляция между малозначимыми объясняющими переменными достаточно высока.

Однако данный признак будет надежным лишь в случае двух объясняющих переменных. При большем их количестве более целесообразным является использование частных коэффициентов корреляции.

3. Высокие частные коэффициенты корреляции.

Частные коэффициенты корреляции определяют силу линейной зависимости между двумя переменными без учета влияния на них других переменных. Однако при изучении многомерных связей в ряде случаев парные коэффициенты корреляции могут давать совершенно неверные представления о характере связи между двумя переменными. Например, между двумя переменными x и y может быть высокий положительный коэффициент корреляции не потому, что одна из них стимулирует изменение другой, а оттого, что обе эти переменные изменяются в одном направлении под влиянием других переменных как учтенных в модели, так и, возможно, неучтенных. Поэтому необходимо измерять действительную силу линейной связи между двумя переменными, очищенную от влияния на рассматриваемую пару переменных других факторов. Коэффициент корреляции между двумя переменными, очищенный от влияния других переменных, называется **частным коэффициентом корреляции**.

Например, при трех объясняющих переменных x_1, x_2, x_3 ($r_{12} = 0,5, r_{13} = 0,5, r_{23} = -0,5$). Тогда частный коэффициент корреляции $r_{1,2,3} = 1$, т. е. при относительно невысоком коэффициенте корреляции

r_{12} частный коэффициент корреляции $r_{1,2,3}$ указывает на высокую зависимость (коллинеарность) между факторами x_1 и x_2 .

4. Сильная вспомогательная (дополнительная) регрессия.

Мультиколлинеарность может иметь место вследствие того, что какая-либо из объясняющих переменных является линейной (или близкой к линейной) комбинацией других объясняющих переменных.

5.2. Методы построения ОЛСМ в условиях мультиколлинеарности факторов

Прежде чем указать основные методы устранения мультиколлинеарности, отметим, что в ряде случаев мультиколлинеарность не является такой уж серьезной проблемой, чтобы прилагать существенные усилия по ее выявлению и устранению. В основном все зависит от целей исследования.

Если основная задача модели – прогноз будущих значений зависимой переменной, то при достаточно большом коэффициенте детерминации R^2 ($>0,9$) наличие мультиколлинеарности обычно не сказывается на прогнозных качествах модели. Хотя это утверждение будет обоснованным лишь в том случае, что и в будущем между коррелированными переменными будут сохраняться те же отношения, что и ранее.

Если же целью исследования является определение степени влияния каждой из объясняющих переменных на зависимую переменную, то наличие мультиколлинеарности, приводящее к увеличению стандартных ошибок, скорее всего, исказит истинные зависимости между переменными. В этой ситуации мультиколлинеарность является серьезной проблемой.

Отметим, что единого метода устранения мультиколлинеарности, годного в любом случае, не существует. Это связано с тем, что причины и последствия мультиколлинеарности неоднозначны и во многом зависят от результатов выборки.

Исключение переменной(ых) из модели.

Простейшим методом устранения мультиколлинеарности является исключение из модели одной или ряда коррелированных переменных.

Однако необходима определенная осмотрительность при применении данного метода. В этой ситуации возможны ошибки спецификации.

Например, при исследовании спроса на некоторое благо в качестве объясняющих переменных можно использовать цену данного блага и

цены заменителей данного блага, которые нередко коррелируют друг с другом. Исключив из модели цены заменителей, мы, скорее всего, допустим ошибку спецификации. Вследствие этого можно получить смещенные оценки и сделать необоснованные выводы.

Таким образом, в прикладных эконометрических моделях желательно не исключать объясняющие переменные до тех пор, пока коллинеарность не станет серьезной проблемой.

Получение дополнительных данных или новой выборки.

Поскольку мультиколлинеарность напрямую зависит от выборки, то, возможно, при другой выборке мультиколлинеарности не будет либо она не будет столь серьезной.

Иногда для уменьшения мультиколлинеарности достаточно увеличить объем выборки.

Например, при использовании ежегодных данных можно перейти к поквартальным данным. Увеличение количества данных сокращает дисперсии коэффициентов регрессии и тем самым увеличивает их статистическую значимость.

Однако получение новой выборки или расширение старой не всегда возможно или связано с серьезными издержками. Кроме того, такой подход может усилить автокорреляцию. Эти проблемы ограничивают возможность использования данного метода.

Изменение спецификации модели.

В ряде случаев проблема мультиколлинеарности может быть решена путем изменения спецификации модели: либо изменяется форма модели, либо добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на зависимую переменную.

Если данный метод имеет основания, то его использование уменьшает сумму квадратов отклонений, тем самым сокращая стандартную ошибку регрессии. Это приводит к уменьшению стандартных ошибок коэффициентов.

Использование предварительной информации о некоторых параметрах.

Иногда при построении модели множественной регрессии можно воспользоваться предварительной информацией, в частности известными значениями некоторых коэффициентов регрессии. Вполне вероятно, что значения коэффициентов, рассчитанные для каких-либо предварительных (обычно более простых) моделей либо для аналогичной модели по ранее полученной выборке, могут быть использованы для разрабатываемой в данный момент модели.

Например, строится регрессия вида: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Предположим, что переменные x_1 и x_2 коррелированы. Для ранее построенной модели парной регрессии $x_2 = \gamma_0 + \gamma_1 x_1 + \nu$ был определен статистически значимый коэффициент γ_1 (для определенности пусть $\gamma_1 = 0,8$), связывающий y с x_1 . Если есть основания считать, что связь между y и x_1 останется неизменной, то можно положить, что $\gamma_1 = \beta_1 = 0,8$. Тогда ЭМ примет вид: $y = \beta_0 + 0,8x_1 + \beta_2 x_2 + \varepsilon \Rightarrow y - 0,8x_1 = \beta_0 + \beta_2 x_2 + \varepsilon$.

Данное уравнение фактически является уравнением парной регрессии, для которого проблема мультиколлинеарности не существует.

Ограниченность использования данного метода обусловлена тем, что, *во-первых*, получение предварительной информации нередко затруднительно, а *во-вторых*, вероятность того, что выделенный коэффициент регрессии будет одним и тем же для различных моделей, невысока.

Преобразование переменных.

В ряде случаев минимизировать либо вообще устранить проблему мультиколлинеарности можно с помощью преобразования переменных.

Например, пусть эмпирическое уравнение регрессии имеет вид: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$, причем x_1 и x_2 – коррелированные переменные. В этой ситуации можно попытаться определять регрессионные зависимости относительных величин: $\frac{\hat{y}}{x_1} = b_0 + b_1 \frac{x_2}{x_1}$, $\frac{\hat{y}}{x_2} = b_0 + b_1 \frac{x_1}{x_2}$.

Вполне вероятно, что в этих моделях, проблема мультиколлинеарности будет отсутствовать. Возможны и другие преобразования, близкие по своей к вышеописанным. Например, если в уравнении рассматриваются взаимосвязи номинальных экономических показателей, то для снижения мультиколлинеарности можно попытаться перейти к реальным показателям и т. п.

Каскадный корреляционный анализ.

Чтобы избежать искажения коэффициентов регрессии в корреляционной модели с мультиколлинеарными факторами, используется каскадный корреляционный анализ.

Сущность каскадного корреляционного анализа заключается в следующем.

1. Как обычно, в соответствии с изложенным выше, выбираем ре-

зультативный и факторные показатели, проверяем информацию столбцов на достоверность.

2. Выясняем пары факторов, тесно связанных друг с другом, т. е. коррелируемых (например, в корреляционной модели формирования стоимости валовой продукции – основные производственные и оборотные фонды).

3. Определяем, какие из факторов тесно связанных пар являются ведущими (определяющими). Эти определяющие факторы назовем промежуточными результативными.

4. Строим парную корреляционную модель взаимосвязи каждой пары факторов, например: $y_{x_2} = a_0 + a_1 x_1$, где y_{x_2} – стоимость оборотных фондов; x_1 – стоимость основных производственных фондов.

При этом рассчитываем все остальные характеристики.

5. Рассчитаем разность фактических и расчетных значений фактора, тесно связанного с другим или другими: $x_2 - y_{x_2} = \Delta x_2$.

В корреляционной модели вместо фактора x_2 ставим столбец Δx_2 , определяющий величину отклонения фактического значения фактора от среднего уровня, и считаем параметры модели. В этом случае коэффициенты регрессии при Δx_2 покажут влияние на результативный показатель нового фактора при его отклонении от среднего уровня. В этом случае удастся избежать искажения, имеющего место в корреляционной модели с тесно коррелируемыми факторами.

6. МОДЕЛИ И МЕТОДЫ АНАЛИЗА СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

6.1. Общие сведения о временных рядах и их классификация.

6.2. Стационарный временной ряд и его характеристики.

6.3. Определение и свойства модели авторегрессии $AR(p)$, $AR(1)$, модели скользящего среднего $MA(m)$, $MA(1)$. Модель $ARMA(p, m)$: свойства, методы построения и тестирования.

6.1. Общие сведения о временных рядах и их классификация

Общие понятия о временных рядах.

Временным рядом называют последовательность наблюдений, обычно упорядоченную во времени (хотя возможно упорядочение и по какому-либо другому параметру). Основной чертой, выделяющей ана-

лиз временных рядов среди других видов статистического анализа, является существенность порядка, в котором производятся наблюдения.

В эконометрии принято моделировать временной ряд как **случайный процесс**, называемый также **стохастическим процессом**, под которым понимается статистическое явление, развивающееся во времени согласно законам теории вероятностей. Случайный процесс – это случайная последовательность.

Различают два вида временных рядов:

- 1) непрерывные;
- 2) дискретные.

Измерение некоторых величин (температуры, напряжения и т. д.) производится непрерывно, по крайней мере, теоретически. При этом наблюдения можно фиксировать в виде графика. Но даже в том случае, когда изучаемые величины регистрируются (или могут регистрироваться) непрерывно, практически при их обработке используются только те значения, которые соответствуют дискретному множеству моментов времени.

Если время фиксируется дискретно (т. е. через фиксированный интервал времени), то временной ряд **дискретен** (y_1, y_2, \dots, y_T , так что интервал $(t, t+1)$ является постоянным). В дальнейшем мы будем иметь дело только с дискретными временными рядами.

Временные ряды подразделяются:

- 1) на детерминированные;
- 2) случайные.

Детерминированным называют процесс, который принимает заданное значение с вероятностью 1. Когда же мы будем говорить о случайном процессе и случайном временном ряде, то, как правило, будем подразумевать, что он существенно случаен, т. е. не является детерминированным.

Временные ряды подразделяются:

- 1) на стационарные;
- 2) нестационарные.

Временной ряд является стационарным, если он находится в определенном смысле в статистическом равновесии, т. е. его свойства с вероятностной точки зрения не зависят от времени. Процесс **нестационарен**, если эти условия нарушаются.

Итак, при построении временного ряда его разделяют на составляющие части:

$$y(t) = f(t) + s(t) + u(t) + e(t), \quad (6.1)$$

где $f(t)$ – функция тренда;

$s(t)$ – сезонная компонента;

$u(t)$ – циклическая компонента;

$e(t)$ – остаточная компонента. Охарактеризуем их более подробно.

- *Тренд* характеризуют долговременную тенденцию развития некоторого явления. При этом она выражается некоторой монотонной функцией. В качестве примера трендов можно указать изменение демографических характеристик, рост экономических показателей, рост потребления.

- *Сезонная компонента* характеризует воздействие факторов, возникающих с определенной периодичностью. Особенностью является то, что их действие заканчивается в течение года. Например, загруженность трассы в течение суток, повышение спроса на товары для школьников в конце августа.

- *Циклическая компонента* – это функция, описывающая явление, действующее с длительным периодом. Особенностью является то, что для выявления циклической компоненты обычно недостаточно только наблюдаемых данных, а требуется анализ общей, экономической, социальной и даже исторической ситуации. Например, демографические ямы.

- *Остаточная компонента* выражает воздействие случайных факторов, и при изучении этой компоненты требуется изучение статистического и вероятностного анализа. В зависимости от характера выделяют: «белый шум», авторегрессию, скользящее среднее и смешанную случайную компоненту.

Временной ряд можно считать состоящим из двух частей:

Временной ряд						
Детерминированная составляющая			Случайная составляющая $e(t)$			
Тренд $f(t)$	Циклическая компонента $u(t)$	Сезонная компонента $s(t)$	«Белый шум»	Авторегрессия	Скользящее среднее	Смешанная

На рис. 6.1 показан график временного ряда, на котором прослеживаются все три составляющие.

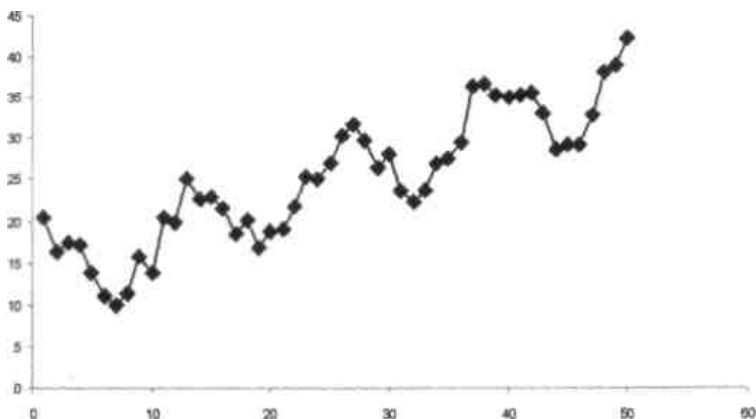


Рис. 6.1. Временной ряд

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется *аддитивной моделью временного ряда*. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется *мультипликативной моделью временного ряда*. Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент, с тем чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

Важной характеристикой временного ряда является **автокорреляция**.

При наличии тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих значений. Корреляционную зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда*.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени. Коэффициент автокорреляции порядка определяется как коэффициент корреляции между рядами y_t , y_{t-1} :

$$r_{\tau} = \frac{\sum_{t=\tau+1}^n (y_t - \overline{y_{1\tau}})(y_{t-\tau} - \overline{y_{2\tau}})}{\sqrt{\sum_{t=\tau+1}^n (y_t - \overline{y_{1\tau}})^2 \sum_{t=\tau+1}^n (y_{t-\tau} - \overline{y_{2\tau}})^2}}, \quad (6.2)$$

где τ – величина сдвига, называемая *лагом*, определяет порядок коэф-

фициента автокорреляции, $\overline{y_{1\tau}} = \frac{\sum_{t=\tau+1}^n (y_t)}{n - \tau}$, $\overline{y_{2\tau}} = \frac{\sum_{t=\tau+1}^n (y_{t-\tau})}{n - \tau}$.

Коэффициенты автокорреляции разных порядков принято обозначать как r_1, r_2, \dots, r_n , где 1, 2, ..., n указывают на номер порядка коэффициента автокорреляции.

Отметим два важных свойства коэффициента автокорреляции.

Во-первых, он строится по аналогии с линейным коэффициентом корреляции и, таким образом, характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

Во-вторых, по знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом они могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции первого, второго и более высоких порядков называется ***автокорреляционной функцией (АКФ) временного ряда***.

График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется ***коррелограммой*** (рис. 6.2).

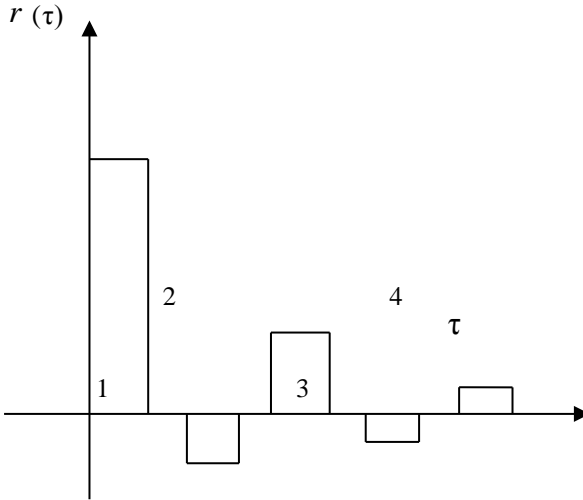


Рис. 6.2. Коррелограмма

Автокорреляционную функцию обычно используют для выявления во временном ряду наличия или отсутствия трендовой и сезонной компонент:

- если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию;
- если наиболее высоким оказался коэффициент автокорреляции порядка $\tau > 1$, то ряд содержит сезонные колебания с периодом τ ;
- если ни один из коэффициентов автокорреляции не является значимым, то либо ряд не содержит тенденции и сезонных колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

6.2. Стационарный временной ряд и его характеристики

Стационарные временные ряды.

Линейные модели временных рядов применяются, как правило, для описания стационарных процессов:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_\tau y_{t-\tau} + \varepsilon_t, \quad t = \overline{1, T},$$

где $y_t, y_{t-1}, \dots, y_{t-\tau}$ – значения переменной в соответствующие моменты времени;

$\alpha_1, \dots, \alpha_\tau$ – параметры модели или коэффициенты модели;

ε_t – случайная ошибка или «белый шум».

На практике при изучении случайных процессов говорят о *слабой стационарности* и *строгой стационарности процессов*.

Ряд называется **строго стационарным** (или *стационарным в узком смысле*), если его свойства не меняются при изменении начала отсчета времени.

Часто используется понятие **слабой стационарности** (или *стационарности в широком смысле*), которое состоит в том, что среднее, дисперсия и ковариации y_t не зависят от момента времени t :

$$M(y_t) = \mu, D(y_t) = \gamma_0, \text{cov}(y_t, y_{t-\tau}) = \gamma_\tau.$$

В дальнейшем под стационарностью будем понимать слабую стационарность. Временной ряд, не удовлетворяющий перечисленным выше свойствам, называется *нестационарным* временным рядом.

Стационарный процесс будет характеризоваться такими свойствами, как:

1) *постоянное математическое ожидание* стационарного ряда, т. е. среднее значение временного ряда, вокруг которого изменяются уровни, является величиной неизменной: $M(y_t) = \mu = \text{const}$, где

$$\mu = \frac{1}{T} \sum_{t=1}^T y_t \quad \text{– оценка среднего значения;}$$

2) *постоянная дисперсия* стационарного ряда, определяющая размах его колебаний относительно среднего значения: $D(y) = M(y_t - \mu)^2 = \sigma^2(y) = \text{const}$; $\sigma^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \mu)^2$ – оценка дисперсии;

3) *постоянная автоковариация* стационарного ряда. Для стационарных рядов автоковариация зависит только от величины t , поэтому $\gamma(0) = D(y) = \sigma^2(y)$.

Автоковариационная функция $\gamma(\tau)$. Значения автоковариационной функции статистически оцениваются по имеющимся наблюдениям временного ряда по формуле

$$\gamma(\tau) = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (y_t - \mu)(y_{t-\tau} - \mu), \quad \text{где } \tau = 1, \dots, T-1, \quad (6.3)$$

или

$$\gamma(\tau) = \text{cov}(y_t, y_{t-\tau}).$$

Очевидно, значение автоковариационной функции при $\tau = 0$ есть не что иное, как дисперсия временного ряда, и соответственно

$$\hat{\gamma}(0) = \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2;$$

4) *постоянство коэффициентов автокорреляции* стационарного ряда, т. е.

$$r(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}, \tau = 1, \dots, T-1.$$

Автокорреляционная функция $r(\tau)$ при этом примет такой вид:

$$r(\tau) = \frac{\text{cov}(y_t, y_{t-\tau})}{\sqrt{\text{var}(y_t) \text{var}(y_{t-1})}}. \quad (6.4)$$

Частным случаем стационарных временных рядов является случайный процесс, если $r(\tau) = 0$ при всех τ , то члены ряда (остатки ε_t) некоррелированы. Процесс такого вида называют *белым шумом*.

Определение. Последовательность случайных величин ε_t называется *белым шумом*, если $\forall t = \overline{1, n}, M\varepsilon_t = 0, D\varepsilon_t = \sigma^2 = \text{const}$ и $\text{cov}(\varepsilon_t, \varepsilon_{t+k}) = 0, \forall k > 1$.

Таким образом, «белый шум» – это последовательность некоррелированных случайных величин, одинаково распределенных и имеющих нулевое математическое ожидание и постоянную дисперсию σ^2 . Иногда добавляется следующее требование: $\varepsilon_t \propto N(0, \sigma^2)$, которое означает, что величины ε_t распределены по нормальному закону с математическим ожиданием, равным нулю, и дисперсией, равной σ^2 . В этом случае говорят, что шум – гауссовский «белый шум»;

5) *частная автокорреляционная функция* $\rho_{\text{част}}(\tau)$. Наряду с автокорреляционной функцией при исследовании стационарных временных рядов рассматривается частная автокорреляционная функция $r_{\text{част}}(\tau)$ (ЧАКФ). С помощью ЧАКФ измеряется корреляция между уровнями ряда y_t и $y_{t-\tau}$, разделенными τ -временными тактами, при исключении

влияния на эту взаимосвязь всех промежуточных уровней ряда: $y_{i-1}, y_{i-2}, \dots, y_{i-\tau+1}$.

Частная автокорреляция 1-го порядка может быть подсчитана с использованием соотношения

$$r_{\text{част}}(2) = r(y_i, y_{i-2} | y_{i-1} = \bar{y}) = \frac{r(2) - r^2(1)}{1 - r^2(1)}. \quad (6.5)$$

ЧАКФ также должна быстро убывать для стационарного процесса.

Проверка на стационарность.

Первое, что следует сделать, – построить график временного ряда.

График может содержать очевидный на глаз тренд или сезонную компоненту. Также возможно, что разброс наблюдений возрастает или убывает со временем. Это может служить указанием на зависимость среднего или дисперсии от времени, т. е. ряд будет, скорее всего, нестационарным.

Второе – построить график выборочной АКФ или коррелограммы $r(\tau)$, являющейся статистической оценкой $r(\tau)$.

Коррелограмма стационарного временного ряда быстро убывает с ростом τ после нескольких первых значений. Если график убывает достаточно медленно, есть основание предположить нестационарность ряда.

Кроме выборочной АКФ можно также построить график выборочной ЧАКФ $r_{\text{ч}}(\tau)$, которая также должна быстро убывать для стационарного процесса.

Третье – можно использовать тесты на наличие единичного корня.

Большое значение в анализе временных рядов имеют стационарные временные ряды, вероятностные свойства которых не изменяются во времени. Это объясняется тем, что многие временные ряды могут быть приведены к стационарному ряду после выделения и удаления из них тренда, сезонной компоненты или взятия разностей. Как правило, ряд ошибок является стационарным рядом.

Наиболее распространенными моделями стационарных рядов являются *модели авторегрессии* и *модели скользящего среднего*.

6.3. Определение и свойства модели авторегрессии $AR(p)$, $AR(1)$, модели скользящего среднего $MA(m)$, $MA(1)$. Модель $ARMA(p, m)$: свойства, методы построения и тестирования

Модель авторегрессии порядка p – $AR(p)$.

Пусть имеется временной ряд y_1, y_2, \dots, y_n , или $y_t, t=1, 2, \dots, n$, где y_t – текущее значение уровня.

Основное предположение состоит в том, что текущее значение уравнения ряда y_t является линейной комбинацией p предыдущих значений и случайной ошибки.

Общая модель авторегрессии:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t, \quad (6.6)$$

где $p \leq t$ и $\alpha_1, \dots, \alpha_p$ – параметры модели или коэффициенты модели;

ε_t – случайная ошибка или «белый шум».

Введем оператор лага или обратного действия L : $L(y_t) = y_{t-1}$;
 $L^2(y_t) = y_{t-2}$; $L^p(y_t) = y_{t-p}$.

Тогда можно представить модель в более коротком виде:

$$(1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p) = \varepsilon_t \text{ или } \Phi(L)y_t = \varepsilon_t,$$

где L – оператор сдвига, т. е. преобразование ряда, смещающее его на один временной такт;

$\Phi(L)$ – полином от оператора сдвига.

Удобным и полезным инструментом для изучения процессов авторегрессии является *характеристический многочлен* (характеристический полином):

$$\Phi(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0$$

и связанное с ним *характеристическое уравнение*:

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0.$$

Для выполнения условия стационарности все корни многочлена $\Phi(z)$ должны лежать вне единичного круга, т. е. все корни z соответ-

ствующего характеристического уравнения должны быть по модулю больше 1 и различны.

Например, рассмотрим процесс авторегрессии $AR(2)$: $y_t = -0,9y_{t-1} - 0,2y_{t-2} + \varepsilon_t$.

Характеристическое уравнение принимает вид: $1 + 0,9z + 0,2z^2 = 0$, или $z^2 + 4,5z + 5 = 0$. При этом его корни $z_1 = -2,5$ и $z_2 = -2$ ($z_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$) по абсолютной величине больше 1, процесс – стационарный.

Простейшим примером является модель $AR(1)$, или *марковский процесс*:

$$y_t = \alpha y_{t-1} + \varepsilon_t \quad (6.7)$$

Условие стационарности ряда для $AR(1)$ определяется условием $|\alpha| < 1$, или, что то же самое, корень характеристического уравнения $1 - \alpha z = 0$ должен быть по абсолютной величине больше $|z| > 1$.

Процессы, у которых $|z| = 1$, называются *процессами единичного корня* и являются *нестационарными* (рис. 6.3).

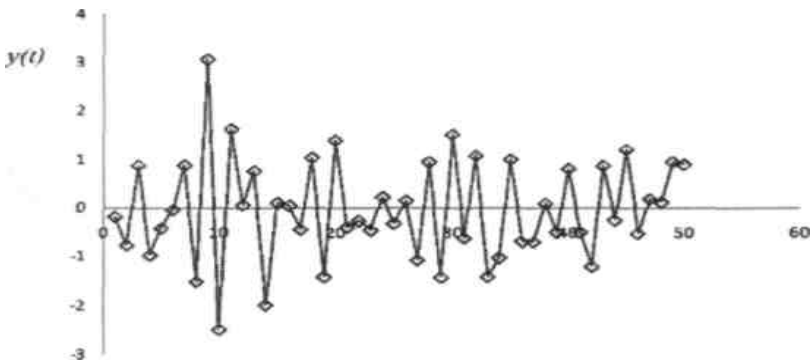


Рис. 6.3. Процесс авторегрессии первого порядка $AR(1)$

Таким образом, механизм порождения последовательных наблюдений, заданный соотношением (6.3), порождает стационарный временной ряд.

В общем случае для процесса $AR(p)$ вытекают следующие *практические рекомендации по их идентификации*:

- значения коэффициентов АКФ экспоненциально затухают (либо монотонно, либо попеременно меняя знак) (рис. 6.4);

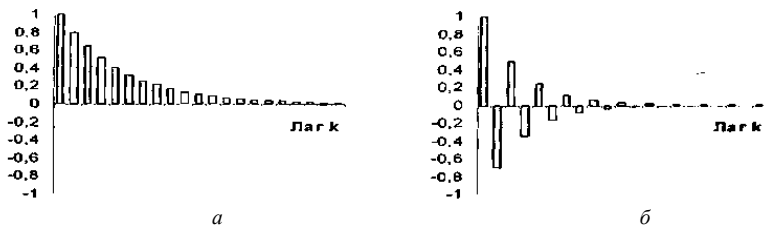


Рис. 6.4. Коррелограмма процесса $AR(1)$:

$$a - \alpha_1 > 0; \text{б} - \alpha_1 < 0$$

- значения коэффициентов ЧАКФ имеют выбросы (пики) на первых p -лагах, а значения коэффициентов для лагов, больших порядка авторегрессии, статистически незначимы (рис. 6.5).

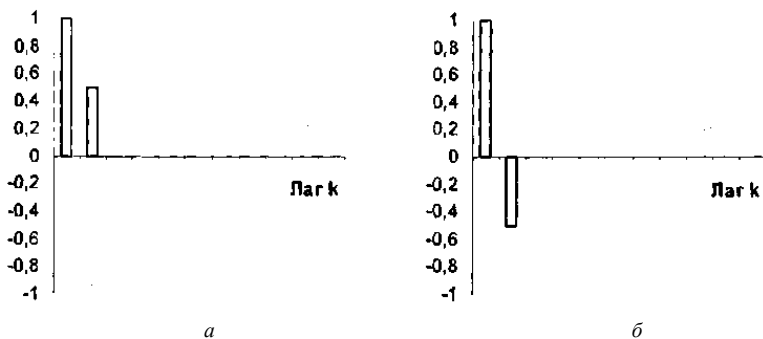


Рис. 6.5. Частная автокорреляционная функция процесса $AR(1)$:

$$a - \alpha_1 > 0; \text{б} - \alpha_1 < 0$$

Модели скользящего среднего $MA(m)$.

Эти модели строят на основании предположения о том, что текущее значение уровня ряда представляется в виде линейной комбинации текущей и прошлых значений ошибки, т. е.

$$y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_m \varepsilon_{t-m}, \quad (6.8)$$

где $\beta_1, \beta_2, \dots, \beta_m$ – параметры модели;

ε_t – процесс белого шума.

Или в более короткой форме: $y_t = \hat{\Phi}(L)\varepsilon_t$.

Название «скользящее среднее» объясняется тем, что текущее значение случайного процесса определяется взвешенным средним m предыдущих значений белого шума. Процедуру скользящего среднего часто используют для того, чтобы сгладить данные, которые сильно колеблются.

Модель скользящего среднего $MA(1)$.

Она представляется следующим выражением:

$$y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1}. \quad (6.9)$$

В общем случае для процесса $MA(q)$ вытекают следующие *практические рекомендации* по их идентификации:

- автокорреляционная функция имеет выбросы (пики) на первых q -лагах, а остальные значения статистически незначимы (рис. 6.6);

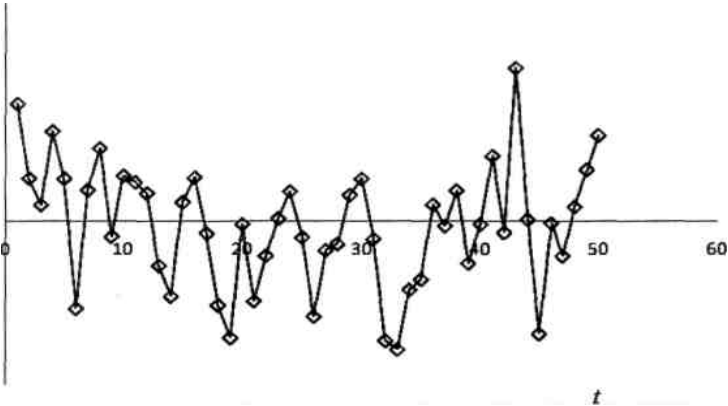


Рис. 6.6. Процесс скользящего среднего $MA(1)$

- частная автокорреляционная функция экспоненциально затухает (либо монотонно, либо осциллируя, т. е. меняя знак) (рис. 6.7).

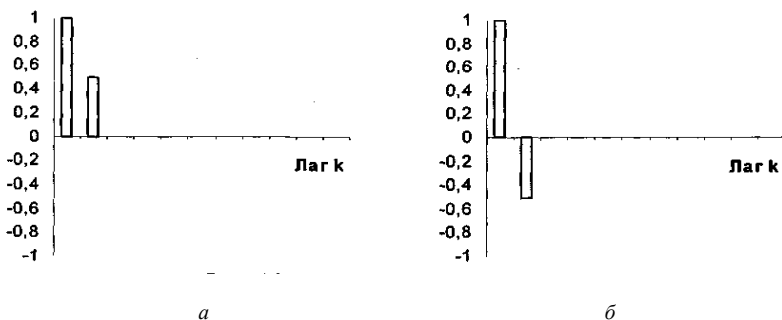


Рис. 6.7. Коррелограмма процесса $MA(1)$:
 $a - \beta_1 < 0$; $\delta - \beta_1 > 0$

Модели $ARMA(p, m)$ – модели авторегрессии-скользящего среднего.

Эти модели основаны на предположении о том, что текущий уровень ряда y_t является линейной комбинацией k своих предыдущих уровней и m своих предыдущих ошибок. При идентификации модели $ARMA(k, m)$ пользуются тем, что их автокорреляционные функции затухают плавно по экспоненте или синусоиде (рис. 6.8).

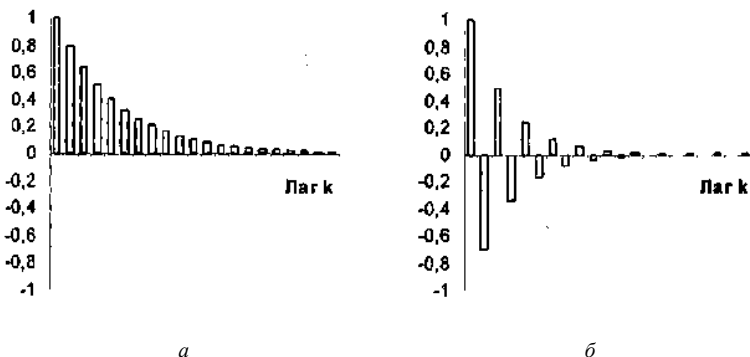


Рис. 6.8. Частная автокорреляционная функция процесса $MA(1)$:
 $a - \beta_1 < 0$; $\delta - \beta_1 > 0$

Общий вид модели:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_m \varepsilon_{t-m}, \quad (6.10)$$

где $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_m$ – коэффициенты модели;

p – порядок авторегрессии;

m – порядок скользящего среднего.

Или в более короткой форме: $\Phi(L)y_t = \hat{\Phi}(L)\varepsilon_t$.

Заметим, что модель (6.10) может быть преобразована в модель авторегрессии $AP(k)$:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_k y_{t-p} + \zeta_t, \quad (6.11)$$

где ошибка ζ_t удовлетворяет свойствам процесса скользящего среднего порядка от порядка m ; $\zeta_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_m \varepsilon_{t-m}$ либо в модель скользящего среднего – $MA(m)$:

$$\begin{aligned} y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} - \dots - \alpha_k y_{t-p} &= \\ &= \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2}, \dots, \beta_m \varepsilon_{t-m}. \end{aligned} \quad (6.12)$$

Рассмотрим процесс $ARMA(1, 1)$:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t - \beta_1 \varepsilon_{t-1}, \quad (6.13)$$

или через оператор сдвига $(1 - \alpha L)y_t = (1 - \beta L)\varepsilon_t$.

Процесс стационарен, если $|\alpha_1| < 1$, и обратим, если $|\beta_1| < 1$ (рис. 6.9).

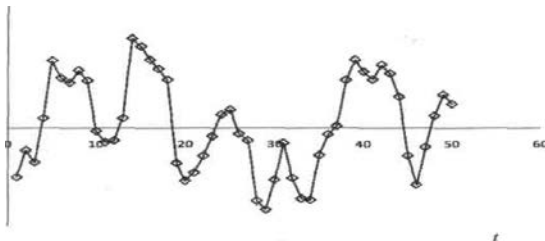


Рис. 6.9. Процесс авторегрессии-скользящего среднего $ARMA(1, 1)$

Обычно число параметров p или q не бывает больше двух. Для процесса $ARMA(p, q)$ вытекают следующие практические рекомендации по их идентификации:

- $ARMA(1, 0)$: АКФ экспоненциально затухает, ЧАКФ имеет выброс на лаге 1;
- $ARMA(2, 0)$: АКФ имеет форму затухающей синусоидальной волны или экспоненциально затухает, ЧАКФ имеет выбросы на лагах 1 и 2;
- $ARMA(0, 1)$: АКФ имеет выброс на лаге 1, ЧАКФ экспоненциально затухает;
- $ARMA(0, 2)$: АКФ имеет выбросы для лагов 1 и 2, ЧАКФ имеет форму затухающей синусоидальной волны или экспоненциально затухает;
- $ARMA(1, 1)$: АКФ экспоненциально затухает от значения $r(1)$, ЧАКФ экспоненциально убывает от значения $r_t(1)$.

Из рассмотренных соотношений вытекает важный вывод: на практике можно подобрать модель с минимальным числом параметров, которая описывает временной ряд y_t , являющийся стационарным процессом второго порядка, не «хуже», чем другие варианты моделей с большим числом параметров. Обычно понятие «не хуже» связывается с минимальной дисперсией модели и отсутствием автокорреляции в ряду ее ошибки.

Практическая ценность этого вывода состоит в следующем. При построении моделей временных рядов нужно стремиться к минимизации числа их параметров, а следовательно, и порядка самой модели. Дело в том, что параметры таких моделей оцениваются на основе коэффициентов автокорреляции исходного процесса y_t . С увеличением порядка модели для определения значений ее параметров необходимо использовать в качестве исходных данных и большее число выборочных коэффициентов автокорреляции (с большими номерами). Точность их оценки с ростом сдвига падает, да и их абсолютные значения либо стремятся к нулю, либо попадают в область повышенной неопределенности. Из-за этого снижается надежность оценок коэффициентов моделей временных рядов высоких порядков, как и качество самих моделей. Все это и заставляет эконометриков искать для описания реальных процессов модели временных рядов с минимальным числом параметров.

7. МОДЕЛИ И МЕТОДЫ АНАЛИЗА НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

7.1. Модели и методы анализа нестационарных временных рядов.

7.2. Определение и свойства модели *ARIMA*, построение, тестирование и прогнозирование.

7.3. Модели временных рядов с условной гетероскедастичностью. Определение и свойства моделей *ARCH* и *GARCH*.

7.1. Модели и методы анализа нестационарных временных рядов

Если временной ряд содержит, например, некоторый тренд, то требование постоянства дисперсии, среднего и ковариации нарушается, и мы имеем дело с нестационарным процессом, или процессом случайного блуждания. На практике временной ряд y_t можно субъективно определить как нестационарный при помощи графика временного ряда и его коррелограмм. Если на них обнаруживаются: тренд, детерминистическая периодичность, гетероскедастичность, изменяющаяся автокорреляция, то имеются веские причины предполагать, что лежащий в основе процесс является нестационарным. Теоретически он нестационарен, если среднее или дисперсия, или ковариация случайного процесса, сгенерировавшего этот временной ряд, изменяются во времени.

Как правило, временные ряды, характеризующие экономические явления, отличаются нестационарностью. Это связано с некоторыми свойствами экономических временных рядов, прежде всего с *наличием тренда*. Очевидно, что при наличии трендовой компоненты сложно утверждать, что среднее значение ряда, а также его дисперсия и автоковариация не зависят от времени, следовательно, ряд нестационарен. Временные ряды могут иметь как строго возрастающий (убывающий) тренд, так и заметные колебания на фоне общего тренда. Подобное поведение характерно для показателей ВВП, а также для показателей инфляции и процентной ставки (рис. 7.1).

Для некоторых нестационарных временных рядов характерно *случайное блуждание*. Обычно такие временные ряды не вызывают тенденции ни к возрастанию, ни к убыванию. Временной ряд может возрастать или убывать со временем и не сохранять среднего значения в долгосрочном периоде.

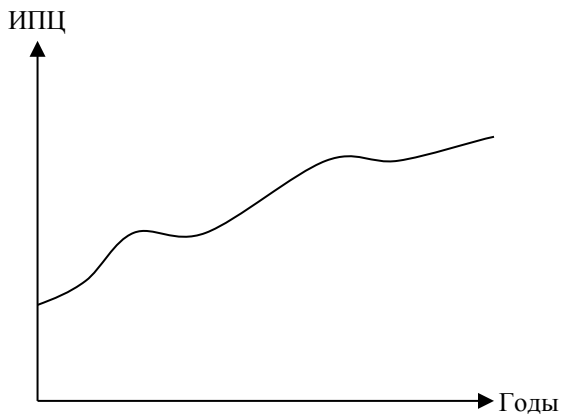


Рис. 7.1. Динамика индекса потребительских цен

Еще одной из причин, вызывающих нестационарность временных рядов, является **высокая инерционность внезапного воздействия (шока)** на временной ряд. Во время экономического спада или бума основные макроэкономические показатели претерпевают сильные изменения и остаются на новом уровне в течение длительного промежутка времени, не возвращаясь к своему прежнему положению. Типичным примером может служить динамика обменного курса доллара во время финансового кризиса.

Относительно длинные временные ряды, характеризующие, например, процессы инфляции или уровень инвестиций, в некоторых случаях можно охарактеризовать как условно гетероскедастичные. Это означает, что в долгосрочном периоде (на протяжении нескольких десятков лет) дисперсия ряда постоянна, но в рамках данного периода имеются более короткие отрезки времени (продолжительностью в несколько лет), на протяжении которых дисперсия явления относительно высока.

Идентификация рядов, основанная на проверке постоянства среднего, дисперсии и ковариации, невозможна, так как априори структура ряда неизвестна.

Для получения критерия, который можно было бы использовать для выявления нестационарности рядов, рассмотрим авторегрессионный процесс y_t первого порядка:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t. \quad (7.1)$$

Несложно проверить, что при $|\alpha_1| < 1$ условия выполняются, а при $|\alpha_1| = 1$ – не выполняются, т. е. в первом случае можно говорить о стационарном, а во втором случае – о нестационарном процессе y_t , поэтому нестационарные процессы называют также *процессами единичного корня*.

Между стационарными и нестационарными временными рядами имеется существенное отличие. Одновременное шоковое воздействие на стационарный ряд носит временный характер. Со временем эффект рассеивается и значения временного ряда возвращаются к своему долгосрочному среднему значению. Следовательно, долгосрочный прогноз стационарного ряда сходится к безусловному среднему.

Нестационарные ряды обязательно имеют постоянную компоненту, среднее и (или) дисперсия зависят от времени. Перечисленные ниже **свойства** помогут идентифицировать нестационарные временные ряды.

1. В долгосрочном периоде не существует постоянного среднего значения, к которому возвращаются значения временного ряда.
2. Дисперсия зависит от времени и по мере увеличения времени растет до бесконечности.
3. Теоретическая автокорреляция не сокращается, но для наблюдений, ограниченных некоторыми пределами, медленно затухает.

В основу тестов на идентификацию временных рядов положена проверка условия равенства или неравенства параметра α_1 , из уравнения (7.1), единице. Это так называемые тесты единичного корня.

Метод разностей и интегрируемость.

С одной стороны, большинство экономических временных рядов нестационарны, а с другой стороны, многие методы и модели основаны на предположении о стационарности временных рядов.

Во многих случаях взятие разностей временных рядов позволяет получить стационарные временные ряды.

Первые разности стохастического процесса имеют такой вид:

$$(1 - L)y_t = \Delta y_t = y_t - y_{t-1}. \quad (7.2)$$

Если первые разности ряда y стационарны, то ряд y_t называется интегрируемым первого порядка.

В противном случае дальнейшее взятие разностей приведет ко вторым разностям:

$$(1 - L)^2 y_t = \Delta^2 y_t = \Delta y_t - \Delta y_{t-1}. \quad (7.3)$$

Если этот ряд стационарен, то ряд y_t называется интегрируемым второго порядка. Если мы получаем первый стационарный ряд после k -кратного взятия разностей, процесс называется интегрируемым k -го порядка. Временной ряд, сгенерированный случайным процессом, интегрируемым k -го порядка, также называется интегрируемым k -го порядка.

7.2. Определение и свойства модели ARIMA, построение, тестирование и прогнозирование

Таким образом, нестационарные временные ряды могут быть сведены к стационарным с помощью оператора последовательной разности.

Пусть временной ряд y_t после применения к нему d раз оператора последовательной разности стал стационарным рядом $\Delta^d y_t$, удовлетворяющим ARMA(p, q)-модели. В этом случае процесс y_t принято называть **интегрированным процессом авторегрессии и скользящего среднего**, или **ARIMA(p, d, q)**. В специальной литературе это модель **Бокса – Дженкинса**.

Модель ARIMA обладает тремя параметрами: p – порядок авторегрессии AR; d – порядок последовательных разностей уровней временных рядов, обеспечивающий стационарность ряда; q – порядок скользящей средней MA. Из модели для ряда Δy_t можно получить модель для исходного ряда y_t , используя соотношение $y_t = y_{t-1} + \Delta y_t$.

Например, модель ARMA (2, 1): $y_t = 1,2y_{t-1} - 0,2y_{t-2} + \varepsilon_t - 0,5\varepsilon_{t-1}$ имеет для AR-части характеристическое уравнение $1 - 1,2z + 0,2z^2 = (1 - 0,2z)(1 - z) = 0$, корни которого $z_1 = 5, z_2 = 1$.

Поскольку один из корней равен 1, то процесс ARMA(2, 1) является нестационарным. Оператор авторегрессии этого процесса можно представить в следующем виде:

$$(1 - 0,2L)(1 - L) = (1 - 0,2L)\Delta.$$

Введем обозначение $\Delta y_t = y_t + \Delta y_{t-1}$. Полученный процесс Δy_t является стационарным процессом $ARMA(1, 1)$, задаваемым уравнением $\Delta y_t = 0,2\Delta y_{t-1} + \varepsilon_t - 0,5\varepsilon_{t-1}$, при этом исходный ряд y_t является рядом $ARIMA(1, 1, 1)$, или проинтегрированным первого порядка $ARMA(1, 1)$ -рядом.

Методология **Бокса – Дженкинса** подбора $ARIMA$ -модели для описания и прогнозирования временного ряда включает следующие этапы:

- 1-й этап: идентификация модели;
- 2-й этап: оценивание модели и проверка ее адекватности;
- 3-й этап: прогнозирование.

1-й этап. Диагностика, т. е. проверка временного ряда на стационарность:

- изучение графика временного ряда;
- тест на единичный корень.

В случае нестационарности – взятие разностей и повтор тестов.

Оценивание диагностических функций, таких, как автокорреляционная, и исследование их графиков.

Выбор типов возможных процессов, сгенерировавших этот временной ряд, так называемая идентификация модели.

В результате должны быть получены три основных параметра: d – порядок интегрируемости, p и q – порядки компонент AR и MA соответственно.

В процессе диагностики параметр d легко определяется как количество взятых разностей, необходимое для получения стационарного процесса. Для экономических временных рядов параметр d обычно равен 1, но иногда он может быть равен 0 или 2.

Более сложен выбор параметров p и q . Обычно при выборе этих параметров полагаются на результаты исследования автокорреляционной функции (АКФ), частной автокорреляционной функции (ЧАКФ). В случае сомнений следует придерживаться правила выбора модели с наименьшим возможным числом параметров.

2-й этап. Оценивание параметров для всех возможных версий модели подходящими статистическими методами, такими как:

- обычный метод наименьших квадратов;
- метод максимального правдоподобия.

3-й этап. Выбор наиболее подходящей модели среди оцененных:

- проверка модели;
- анализ остатков, которые должны иметь свойства «белого шума»;

- рассмотрение модели, наилучшим образом воспроизводящей конкретный временной ряд, и ее наиболее экономичного с точки зрения количества параметров. На основе сравнения критериев выбирается оптимальная модель и на ее основе можно производить прогнозирование.

7.3. Модели временных рядов с условной гетероскедастичностью. Определение и свойства моделей ARCH и GARCH

В 2003 г. американец Роберт Энгле получил Нобелевскую премию за разработку метода анализа экономических временных рядов на основе математической модели с авторегрессионной условной гетероскедастичностью (*ARCH*). Данная модель позволяет прогнозировать тенденции изменения финансовых индексов, таких как ВВП, потребительские цены, процентные ставки, биржевые курсы и др., как на ближайший день, так и на неделю и даже год вперед.

Экономические данные обычно представлены в виде временных рядов, т. е. последовательности наблюдений в хронологическом порядке. Например, величины ВВП за различные годы, цены на товар в различные дни месяца и т. д. Такие ряды можно представить в виде суммы двух компонентов, один из которых изменяется случайным образом, а другой подчиняется определенному закону.

На финансовых рынках случайные отклонения величины от постоянного значения с течением времени (так называемая колеблемость) особенно важны, поскольку стоимость акций, опционов и других финансовых инструментов зависит от уровня их рисков. Колебания могут значительно меняться во времени: периоды сильных изменений сменяются периодами незначительных отклонений. Несмотря на меняющуюся колеблемость, экономисты применяли статистические методы, в которых предполагалось, что колеблемость постоянна.

Но в 1982 г. Роберт Энгле обнаружил, что авторегрессионная гетероскедастическая (т. е. предполагающая изменение во времени) модель очень точно описывает множество временных рядов, встречающихся в экономике. Его метод в настоящее время стоит на вооружении финансовых аналитиков, использующих *ARCH* для оценки стоимости активов и рисков портфельных инвестиций.

Колеблемость обычно измеряется дисперсией σ^2 временного ряда, или стохастического процесса.

Условная гетероскедастичность (УГ) означает, что условная дисперсия ошибки, т. е. дисперсия при условии известной информации, зависит от времени. Она может проявляться несмотря на **общую гомоскедастичность (безусловную)**.

Дисперсия ошибок модели ε_t имеет вид: $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$.

Но условная дисперсия в соответствии с имеющейся на последний момент информацией определяется как $\text{var}(\varepsilon_t | \varepsilon_{t-1})$.

Модели ARCH/GARCH.

1. **ARCH-модель.** В $ARCH(p)$ -модели, предложенной Р. Энгле, предполагается, что дисперсия является линейной функцией квадратов предшествующих значений наблюдаемой величины:

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{n-i}^2. \quad (7.4)$$

Для того чтобы эта величина оставалась положительной с вероятностью 1, требуется $\alpha_0 > 0$, $\alpha_i \geq 0$.

В рамках $ARCH$ -модели стало возможным объяснить такой феномен финансовых временных рядов, как кластерность, который состоит в том, что большие (малые) значения y_n влекут за собой большие (малые) последующие значения, но непредсказуемого знака.

Из функции (7.4) видно, что величины σ_n^2 являются функциями от предшествующих значений y_{n-1}, \dots, y_{n-p} .

Таким образом, суть модели $ARCH$ состоит в том, что если абсолютная величина y_n является большой, то это приводит к повышению условной дисперсии в последующие периоды. В свою очередь, при высокой условной дисперсии более вероятно появление больших (по абсолютной величине) значений y_n . Наоборот, если значения y_n в течение нескольких периодов близки к нулю, то это приводит к понижению условной дисперсии в последующие периоды практически до уровня α_0 . В свою очередь, при низкой условной дисперсии более вероятно появления малых (по абсолютной величине) значений y_n . Таким образом, $ARCH$ -процесс характеризуется инерционностью условной дисперсии (кластеризацией колеблемости).

Эффект кластеризации колеблемости отмечен для многих высокочастотных рядов, таких, как изменение цен акций, валютных курсов, доходности спекулятивных активов.

Самая простая $ARCH$ -модель при $p = 1$, $ARCH(1)$, выглядит так:

$$y_n = \sigma_n \varepsilon_n, \quad \sigma_n^2 = \alpha_0 + \alpha_1 y_{n-1}^2.$$

2. $GARCH$ -модель. Исторически одним из первых обобщений модели $ARCH(p)$ стала обобщенная $ARCH$ -модель, или $GARCH$, предложенная Т. Боллерслевом [5]. Эта модель характеризуется двумя параметрами p и q и обозначается $GARCH(p, q)$. В этой модели, как и $ARCH(p)$ -модели, $y_n = \sigma_n \varepsilon_n$, но относительно формирования колеблемости σ_n предполагается, что

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{n-i}^2 + \sum_{j=1}^q \beta_j \sigma_{n-j}^2.$$

Основным преимуществом $GARCH(p, q)$ -моделей перед $ARCH(p)$ -моделью является то, что при подгонке статистических данных моделям $ARCH(p)$ часто приходится обращаться к слишком большому значению p , в то время как при подгонке $GARCH(p, q)$ -моделям можно ограничиваться лишь небольшими значениями p и q .

В простейшем варианте $GARCH(1, 1)$ имеем:

$$\sigma_n^2 = \alpha_0 + \alpha_1 y_{n-1}^2 + \beta_1 \sigma_{n-1}^2. \quad (7.5)$$

Как и в модели $ARCH$, σ^2 служит условной дисперсией процесса $y_n | y_{n-1}$.

Обобщенная авторефессионная модель с условной гетероскедастичностью $GARCH(p, q)$ описывает процесс, в котором условная дисперсия ошибки находится в зависимости от всей доступной в момент времени t информации.

8. МНОГОМЕРНЫЕ ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ И МЕТОДЫ ИХ ПОСТРОЕНИЯ

8.1. Структурная и приведенная формы уравнений.

8.2. Методы оценивания структурных уравнений.

8.1. Структурная и приведенная формы уравнений

Системный подход в экономическом анализе предполагает сложную структуру взаимосвязей между признаками, когда эффективность деятельности экономического объекта характеризуется несколькими по-

казателями. В этом случае одного регрессионного уравнения может быть недостаточно, и для описания явления или процесса может потребоваться система уравнений и тождеств.

Система одновременных уравнений получила название также структурной формы модели. **Структурной формой модели** (системой одновременных уравнений) называется система уравнений, в каждом из которых аргументы помимо объясняющих переменных могут включать в себя также объясняемые переменные из других уравнений системы.

Уравнения, составляющие исходную модель, называются *структурными уравнениями модели*.

Простейшая структурная форма модели имеет вид:

$$\begin{cases} y_1 = \alpha_1 + \beta_{12}y_2 + \alpha_{11}x_1 + \alpha_{12}x_2 + \varepsilon_1; \\ y_2 = \alpha_2 + \beta_{21}y_1 + \alpha_{21}x_1 + \alpha_{22}x_2 + \varepsilon_2, \end{cases} \quad (8.1)$$

где y , x – зависимая и независимая переменные;

$\varepsilon_1, \varepsilon_2$ – случайные члены;

α, β – параметры модели.

Параметры структурной формы модели называются **структурными коэффициентами**.

Структурная форма модели обычно включает в систему не только уравнения, отражающие взаимосвязи между отдельными переменными, но и уравнения, отражающие тенденцию развития явления, а также разного рода уравнения-тождества. Тождества не содержат каких-либо подлежащих оценке параметров, а также не включают случайный член.

В процессе оценивания параметров одновременных уравнений следует различать эндогенные и экзогенные переменные. Приставки «эндо» и «экзо» означают соответственно «внутреннее» и «внешнее».

Эндогенными считаются переменные, значения которых определяются *внутри* модели и являются зависимыми переменными.

Экзогенными считаются переменные, значения которых определяются *вне* модели и являются независимыми переменными.

В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (лаговые переменные).

Предполагается, что в каждом уравнении экзогенные переменные некоррелированы со случайным членом.

В общем случае эндогенные переменные коррелированы со случайным членом, поэтому применение МНК к структурной форме модели приводит к *смещенным* и *несостоятельным* оценкам структурных коэффициентов.

Для определения структурных коэффициентов структурная форма модели преобразуется в приведенную форму.

Приведенной формой модели называется система уравнений, в каждом из которых эндогенные переменные выражены только через экзогенные переменные и случайные составляющие.

Например, приведенная форма исходной модели имеет вид:

$$\begin{cases} y_1 = \alpha'_1 + \alpha'_{11}x_1 + \alpha'_{12}x_2 + v_1; \\ y_2 = \alpha'_2 + \alpha'_{21}x_1 + \alpha'_{22}x_2 + v_2, \end{cases} \quad (8.2)$$

где α' – параметры приведенной формы;

v_1, v_2 – случайные члены.

Параметры приведенной формы модели называются *коэффициентами приведенной формы* (приведенными коэффициентами).

Коэффициенты приведенной формы оцениваются обычным МНК, поскольку экзогенные переменные некоррелированы со случайным членом.

Оцененные коэффициенты приведенной формы могут быть использованы для оценивания структурных коэффициентов. Такой способ оценивания структурных коэффициентов называется *косвенным МНК*.

Приведенная форма модели аналитически уступает структурной форме модели, так как в ней отсутствуют оценки взаимосвязи между эндогенными переменными.

При переходе от приведенной формы модели к структурной возникает проблема идентификации.

Идентификация – это единственность соответствия между приведенной и структурной формами модели.

Тот или иной структурный коэффициент может либо однозначно выражаться через приведенные коэффициенты, либо иметь несколько разных оценок, либо совсем не выражаться через них.

Структурный коэффициент называется *идентифицируемым*, если его можно вычислить на основе приведенных коэффициентов, причем

точно идентифицируемым, если он единственен, и *сверхидентифицируемым*, если имеет несколько разных оценок; в противном случае он называется *неидентифицируемым*.

Какое-либо структурное уравнение называется идентифицируемым, если идентифицируемы все его коэффициенты. Если хотя бы один структурный коэффициент неидентифицируем, то и все уравнение является неидентифицируемым.

Модель считается идентифицируемой, если каждое ее уравнение идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то вся модель неидентифицируема.

8.2. Методы оценивания структурных уравнений

Рассмотрим различные виды структурных уравнений.

1. Точная идентифицируемость.

Допустим, требуется оценить параметры уравнения функции потребления в простой модели Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t (\text{функция потребления}), \\ Y_t = C_t + I_t (\text{тождество дохода}), \end{cases} \quad (8.3)$$

где C_t , Y_t , I_t – объем потребления, совокупный доход и инвестиции соответственно;

ε_t – случайный член.

Структурный коэффициент β характеризует предельную склонность к потреблению, т. е. из каждой единицы валового внутреннего продукта расходуется β единиц на конечное потребление.

В данной модели C_t , Y_t – эндогенные переменные, а I_t – экзогенная.

Непосредственное оценивание параметров α , β в структурном уравнении функции потребления дает *смещенные* и *несостоятельные* оценки, так как объясняющая переменная Y_t является эндогенной.

Разрешая структурную систему относительно эндогенных переменных, получим приведенную систему:

$$\begin{cases} C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} I_t + \frac{\varepsilon_t}{1-\beta} \\ y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I_t + \frac{\varepsilon_t}{1-\beta} \end{cases} \quad (8.4)$$

В приведенной системе коэффициенты при переменной I_t , равные $MC = \beta / (1 - \beta)$ и $MY = 1 / (1 - \beta)$ – *инвестиционные мультипликаторы потребления и дохода* соответственно. Это значит, что если объем инвестиций возрастает на единицу, то объем потребления увеличится на $\beta / (1 - \beta)$ единиц, а совокупный доход – на $1 / (1 - \beta)$ единиц.

Косвенный метод наименьших квадратов. Уравнение для C_t в приведенной форме можно также представить в следующем виде:

$$C_t = \alpha' + \beta' I_t + \varepsilon'_t, \quad (8.5)$$

где $\alpha' = \frac{\alpha}{1-\beta}$; $\beta' = \frac{\beta}{1-\beta}$; $\varepsilon'_t = \frac{\varepsilon_t}{1-\beta}$.

В этом уравнении экзогенная переменная I_t некоррелирована со случайным членом ε_t , поэтому для оценки параметров (α', β') можно использовать обычный МНК.

Оцененное уравнение $\hat{C}_t = \alpha' + \beta' I_t$, полученное по выборочным данным с помощью МНК, дает *несмещенные* и *состоятельные* оценки параметров.

Поскольку оценки (α, β) структурных коэффициентов $\alpha = \frac{\alpha'}{1-\beta}$,

$\beta' = \frac{\beta}{1-\beta}$ однозначно выражаются через оценки (α', β') приведенных коэффициентов, то структурное уравнение функции потребления является ***точно идентифицируемым***.

Таким образом, для решения точно идентифицируемого уравнения применяется косвенный метод наименьших квадратов (КМНК).

Процедура КМНК производится в несколько этапов.

1. Структурная модель преобразуется в приведенную форму.
2. Для каждого приведенного уравнения обычным МНК оцениваются приведенные коэффициенты.
3. Оценки приведенных коэффициентов преобразуются в оценки параметров структурных уравнений.

2. Сверхидентифицированность. Двухшаговый метод наименьших квадратов (ДМНК).

Рассмотрим следующую простую модель Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t (\text{функция потребления}), \\ Y_t = C_t + I_t + G_t (\text{тождество дохода}), \end{cases} \quad (8.6)$$

где G_t – объем государственных расходов.

В исходной модели C_t , Y_t – эндогенные переменные, а I_t , G_t – экзогенные. Обе экзогенные переменные I_t , G_t не присутствуют в структурном уравнении функции потребления и могут использоваться как инструментальные для эндогенной переменной Y .

Структурное уравнение модели, в которой число экзогенных переменных, которые могут использоваться как инструментальные, больше, чем необходимо, является *сверхидентифицируемым*.

Наилучшим решением в данном случае является применение **двухшагового метода наименьших квадратов (ДМНК)** и построение инструментальной переменной, которая является комбинацией I_t , G_t .

Процедура ДМНК производится в несколько этапов.

1. На основе приведенной формы модели получают для сверхидентифицированного уравнения теоретические (расчетные) значения эндогенных переменных, содержащихся в правой части уравнения.

2. Подставляя теоретические значения эндогенных переменных вместо их фактических значений в сверхидентифицируемое уравнение и применяя обычный МНК, определяют его структурные коэффициенты.

ДМНК можно рассматривать как способ конструирования наилучшей из возможных комбинаций инструментальных переменных в случае, когда в уравнении имеется избыток экзогенных переменных, которые можно использовать как инструментальные для объясняющей эндогенной переменной.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Эконометрика: учебник для магистров / И. И. Елисеева [и др.]; под ред. И. И. Елисеевой. – Москва: Изд-во «Юрайт», 2014. – 453 с.
2. Высшая математика для экономистов: учебник: в 3 т. / И. В. Гайшун [и др.]. – Минск: БГЭУ, 2005. – Т. 2: Теория вероятностей в экономике. Методы оптимизации и экономические модели. – 623 с.
3. Эконометрика: учеб. пособие / А. И. Новиков. – 2-е изд., испр. и доп. – Москва: ИНФРА-М, 2007. – 144 с.
4. Эконометрика и ЭММ: модели множественной регрессии: метод. указания к лабораторным занятиям / И. В. Шафранская. – Горки: БГСХА, 2012. – 56 с.
5. Эконометрика: учебник / И. И. Елисеева [и др.]; под ред. И. И. Елисеевой. – Москва: Проспект, 2010. – 288 с.
6. Грицан, В. Н. Эконометрика: учеб. пособие / В. Н. Грицан. – Москва: Маркетинг, 2001. – 75 с.
7. Дайитбегов, Д. М. Компьютерные технологии анализа данных в эконометрике / Д. М. Дайитбегов. – 2-е изд., испр. и доп. – Москва: ИНФРА-М: Вузовский учебник, 2011. – 578 с.
8. Бородич, С. А. Эконометрика: учеб. пособие / С. А. Бородич. – 3-е изд., стер. – Минск: Новое знание, 2006. – 407 с.
9. Гладилин, А. В. Практикум по эконометрике: учеб. пособие / А. В. Гладилин, А. Н. Герасимов, Е. И. Громов. – Ростов-на-Дону: Феникс, 2011. – 326 с.
10. Ленькова, Р. К. Эконометрика и экономико-математические методы и модели в АПК: учеб. пособие / Р. К. Ленькова, С. П. Старовыборная. – Горки: БГСХА, 2012. – 240 с.
11. Бородич, С. А. Эконометрика. Практикум: учеб. пособие / С. А. Бородич. – Минск: Новое знание; Москва: ИНФРА-М, 2014. – 329 с.
12. Кремер, Н. Ш. Математика для экономистов: от арифметики до эконометрики: учеб.-справ. пособие / Н. Ш. Кремер, Б. А. Путко, И. М. Тришин; под ред. Н. Ш. Кремер; Финанс. ун-т при Правительстве РФ. – 4-е изд., перераб. и доп. – Москва: Юрайт, 2014. – 724 с.

СОДЕРЖАНИЕ

Введение	3
1. Введение в эконометрику (продвинутый уровень)	4
2. Общая линейная статистическая и ее построение с помощью метода наименьших квадратов	8
3. Построение и анализ общей линейной статистической модели в предположении нормального распределения ошибок наблюдения	28
4. Методы построения общей линейной статистической модели при нарушении традиционных предположений относительно ошибок наблюдений	32
5. Построение и анализ общей линейной статистической модели в условиях мультиколлинеарности факторов	46
6. Модели и методы анализа стационарных временных рядов	52
7. Модели и методы анализа нестационарных временных рядов	68
8. Многомерные эконометрические модели и методы их построения	75
Библиографический список	81

Учебное издание

Карачевская Елена Владимировна
Сазонова Светлана Петровна

ЭКОНОМЕТРИКА
(ПРОДВИНУТЫЙ УРОВЕНЬ)

Курс лекций

Редактор *Е. В. Ширалиева*
Технический редактор *Н. Л. Якубовская*

Подписано в печать 28.12.2020. Формат 60×84 ¹/₁₆. Бумага офсетная.
Ризография. Гарнитура «Таймс». Усл. печ. л. 4,88. Уч.-изд. л. 3,85.
Тираж 40 экз. Заказ .

УО «Белорусская государственная сельскохозяйственная академия».
Свидетельство о ГРИИРПИ № 1/52 от 09.10.2013.
Ул. Мичурина, 13, 213407, г. Горки.

Отпечатано в УО «Белорусская государственная сельскохозяйственная академия».
Ул. Мичурина, 5, 213407, г. Горки.