МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ИНСТИТУТ МЕНЕДЖМЕНТА»

И. Г. КИРИН

ЭКОНОМЕТРИКА

Курс лекций

3-е издание

Оренбург
2015

ББК 65 в 631
УДК 330 (075)
Э 40

Рецензенты:

доктор экономических наук, профессор А. П. Тяпухин (ФГБОУ ВПО Оренбургский государственный университет), кандидат экономических наук, доцент М. П. Ситкова (Оренбургский филиал АТ и СО)

Кирин, И. Г.

Э 40 Эконометрика: курс лекций / И. Г. Кирин. 3-е издание, исправленное и дополненное. – Оренбург: ФГБОУ ВПО «ОГИМ», 2015. – 106 с.

ISBN 978-5-9723-0108-9

В курсе лекций изложены общие теоретические положения, которые подготовлены в соответствии с требованиями Государственного образовательного стандарта третьего поколения по экономическим специальностям. Цель курса лекций – подготовка специалистов, владеющих практическими навыками эконометрического моделирования данных, характеризующих состояние и развитие экономики. Рассмотрены линейные и нелинейные модели регрессии, модели стационарных и нестационарных временных рядов, гетероскедастичность, мультиколлениарность, автокорреляция, системы одновременных уравнений.

Для студентов, менеджеров, экономистов, аспирантов, преподавателей экономических вузов, слушателей курсов повышения квалификации и бизнес-школ.

ББК 65 в 631
УДК 330 (075)

ISBN 978-5-9723-0108-9

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Важнейшей задачей курса «Эконометрика» является подготовка специалистов в системе экономического образования, владеющих различными методами сбора, систематизации и анализа сведений, характеризующих экономическое и социальное развитие всех сфер общественной жизни.

Работа современного менеджера, экономиста и других специалистов невозможна без применения приемов и методов эконометрики. Эконометрика для экономических специальностей служит основой для разработки и совершенствования методов экономического анализа, и от степени усвоения этого курса во многом зависит успешность овладения другими экономическими дисциплинами, а следовательно, и умение в дальнейшей практической и научной работе широко пользоваться эконометрическими методами и материалами для решения социальных и экономических проблем.

Настоящий курс лекций охватывает общие начальные элементы эконометрики, и прежде всего – важнейшие направления анализа социально-экономических процессов.

Разработка данного курса лекций в значительной мере связана с тем, что количество часов, отведенных эконометрике в учебных планах, ограничено, а рыночные отношения в России значительно расширились. Поэтому требуется активизация учебного процесса: повышение качества, интенсивности и эффективности преподавания, усиление самостоятельной работы студентов на протяжении всего учебного года, внедрение более эффективных мер контроля за ходом освоения студентами учебного материала и за их успеваемостью. Все это легло в основу изложенных в курсе лекций основных вопросов эконометрики и отдельных ее элементов.

В первой части курса (лекция 1, 2, 3) рассматриваются определение эконометрики, ее предмет, цель и задачи, а также линейные модели парной и множественной регрессии.

Вторая часть (лекция 4) посвящена гетероскедастичности, мультиколлениарности и автокорреляции.

В третьей части курса лекций (лекция 5) рассмотрены нелинейные модели регрессии и их линеаризация.

Четвертая часть курса лекций (лекции 6 и 7) посвящена моделям стационарных и нестационарных временных рядов.

Заключительная (пятая) часть курса лекций (лекции 8 и 9) содержит основные сведения по системам линейных одновременных уравнений.

Приводится словарь терминов и понятий и тесты для самоконтроля.

Предлагаемый курс лекций учит обобщать и анализировать статистические данные, проводить различного рода наблюдения, составлять аналитические таблицы и т. д.

Лекция 1

ЭКОНОМЕТРИКА КАК НАУКА

План:

1. Эконометрика как наука.
2. Предмет эконометрики.
3. Задачи эконометрики в области социально-экономических исследований.
4. Типы эконометрических моделей.

1. Эконометрика как наука

Эконометрика относится к относительно недавно появившимся, но быстро развивающимся научным дисциплинам.

Эконометрика (эконометрия) – наука, изучающая конкретные количественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей. Возникла в начале XX в., термин введен норвежским ученым Р. Фришем.

Формально «Эконометрика» означает измерения в экономике. С. А. Айвазян, раскрывая понятие эконометрики, указывает, что «Эконометрика объединяет совокупность методов и моделей, позволяющих придать количественные выражения качественным зависимостям». Но количественный подход к экономике имеет несколько аспектов, и сам по себе ни один из этих аспектов не следует путать с эконометрикой. Таким образом, эконометрика – это ни в коей мере не то же самое, что экономическая статистика. Она также не совпадает и с тем, что мы называем общей экономической теорией, хотя значительная часть этой теории, безусловно, имеет количественный характер. Эконометрику нельзя также рассматривать как синоним применения математики в экономической теории. Опыт показал, что каждая из этих точек зрения, т. е. статистики, экономической теории и математики, является необходимым, но по отдельности не достаточным условием реального понимания количественных отношений современной экономической жизни. Сила заключается в *объединении* этих трех элементов. И именно это объединение составляет эконометрику.

Таким образом, эконометрику можно рассматривать как науку, которая дает количественное выражение взаимосвязей экономических явлений и процессов, как прикладную экономическую дисциплину, изучающую конкретные количественные взаимосвязи экономических объектов и процессов с помощью математико-статистических методов и моделей.

2. Предмет эконометрики

Предметом исследования эконометрики являются экономические явления. Но в отличие от экономической теории, эконометрика акцентирует внимание на количественных, а не на качественных аспектах этих явлений.

Эконометрика обеспечивает непрерывный процесс принятия управленческих решений, дающих возможность реализовать намеченные цели.

Задачей эконометрики является оценка действий, направленных на достижение и повышение экономической эффективности. Соответственно ключевой целью эконометрики является эмпирическое обобщение экономических законов.

Эконометрика есть единство трех составляющих – статистики, экономической теории и математики, наука, которая дает количественное выражение взаимосвязей экономических процессов и явлений.

В эконометрических исследованиях выделяют следующие этапы:

1-й этап – постановочный, на этом этапе формируется цель исследования, набор участвующих в модели экономических переменных.

2-й этап – априорный, на этом этапе проводится анализ сущности изучаемого объекта, формирование и формализация априорной (известной до начала моделирования) информации.

3-й этап – информационный, на этом этапе осуществляется сбор необходимой статистической информации, значений экономических переменных.

4-й этап – спецификация модели, на этом этапе в математической форме выражаются обнаруженные связи и соотношения, устанавливается состав факторных и результативных переменных; формируются исходные предпосылки и ограничения модели.

5-й этап – параметризация, на этом этапе оцениваются параметры (коэффициенты) выбранной зависимости.

6-й этап – идентификация, на этом этапе осуществляется статистический анализ модели и оценка ее параметров.

7-й этап – верификация, на этом этапе проводится проверка адекватности модели, выясняется, насколько удачно решены проблемы спецификации, идентификации модели, какова точность расчетов по данной модели, насколько соответствует построенная модель реальному экономическому явлению.

Цели эконометрического моделирования следующие:

1) прогноз экономических и социально-экономических показателей, характеризующих состояние и развитие анализируемой системы;

2) имитация различных возможных сценариев социально-экономического развития анализируемой системы (многовариантные сценарные расчеты, ситуационное моделирование).

3. Задачи эконометрики в области социально-экономических исследований

Основная задача эконометрики состоит в том, чтобы наполнить эмпирическим содержанием априорные экономические рассуждения.

Экономисты часто интересуются связями между различными величинами, например, между заработной платой человека и уровнем образования. Наиболее важная задача эконометрики состоит в том, чтобы измерить эти связи количественно на основе имеющихся данных при помощи статистических методов, а также соответствующим образом интерпретировать или использовать полученные результаты. Следовательно, эконометрика – это взаимодействие экономической теории, наблюдаемых данных и статистических методов. Именно взаимодействие этих трех составляющих делает эконометрику интересной, многообещающей и, наверное, трудной. Говоря словами докладчика на одном семинаре, происходившем несколько лет назад, эконометрика намного более проста в отсутствие данных.

Эконометрика дает методы экономического анализа, которые объединяют экономическую теорию со статистическими и математическими методами анализа, позволяет улучшить экономические прогнозы и сделать возможным успешное планирование экономической политики.

Эконометрика используется все более широко, несмотря на то, что полученные с ее помощью прогнозы не всегда оказывались достаточно точными.

Эконометрика связана с изучением эмпирических данных статистическими методами; цель этого изучения состоит в проверке гипотез и оценке соотношений, предложенных экономической теорией. В то время как математическая экономика занимается чисто теоретическими аспектами экономического анализа, эконометрика пытается подвергнуть проверке теории, которые уже представлены в явной математической форме.

4. Типы эконометрических моделей

Выделяют три основных типа (класса) эконометрических моделей, которые применяются для анализа и прогноза: модели временных рядов, регрессионные модели с одним уравнением, системы одновременных уравнений.

Модели временных рядов

К этому классу относят модели:

а) тренда:

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

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

ε – случайная (стохастическая) компонента;

б) сезонности:

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

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

в) тренда и сезонности, причем если эти компоненты складываются $y(t) = T(t) + S(t) + \varepsilon$, то такую модель называют аддитивной моделью, а если эти компоненты умножаются $y(t) = T(t) \cdot S(t) \cdot \varepsilon$, то модель называют мультипликативной.

Регрессионные уравнения с одним уравнением

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

$y = f(x_1, x_2 \ldots\ldots x_n, a_1 \ldots\ldots a_n)$,

где $x_1, x_2 \ldots\ldots x_n$ – независимые (объясняющие) переменные;

$a_1, a_2 \ldots\ldots a_n$ – параметры модели.

В зависимости от вида функции модели делятся на линейные и нелинейные. Область применения таких моделей значительно шире, чем моделей временных рядов.

Системы эконометрических уравнений

Эти модели описываются системами уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может, кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы. Таким образом, имеется набор объясняемых переменных, связанных через уравнения системы.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов, – СПб.: Питер, 2004. – 464 с.

3. Айвазян, С. А. Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001. – 300 с.

4. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва: Дело, 1997. – 442 с.

5. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва : Финансы и статистика, 2005. – 324 с.

6. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

Лекция 2

ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

План:

1. Корреляционная связь.
2. Парная регрессия.
3. Метод наименьших квадратов для построения линейной модели.
4. Линейный коэффициент корреляции.
5. Коэффициент детерминации.
6. Проверка значимости и построения доверительных интервалов коэффициентов регрессии.
7. Интерпретация уравнений.
8. Средняя ошибка аппроксимации.

1. Корреляционная связь

В естественных науках большей частью имеют дело со строгими (функциональными) зависимостями, при которых каждому значению одной переменной соответствует единственное значение другой.

В подавляющем большинстве случаев между экономическими переменными таких зависимостей нет. Например, нет строгой зависимости между доходом и потреблением, ценой и спросом, производительностью труда и стажем работы и т. д. Это связано с целым рядом причин и, в частности, с тем, что, во-первых, при анализе влияния одной переменной на другую не учитывается целый ряд других факторов, влияющих на нее; во-вторых, это влияние может быть не прямым, а проявляться через цепочку других факторов; в-третьих, многие такие воздействия носят случайный характер и т. д. Поэтому в экономике говорят *не о функциональных, а о статистических* зависимостях. Нахождение, оценка и анализ таких зависимостей, построение их математических зависимостей и оценка их параметров являются одним из важнейших разделов эконометрики.

В общем случае статистической называют зависимость, при которой изменение одной из величин влечет изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой. Такую статистическую зависимость называют *корреляционной*.

Корреляционная связь — это частный случай статистической связи. При корреляционной связи разным значениям одной переменной соответствуют различные средние значения другой. Корреляционную связь можно выявить только в виде общей тенденции при массовом изучении факторов

Корреляционно-регрессионный анализ позволяет установить тесноту, направление связи и форму этой связи, т. е. ее аналитическое выражение. Задача корреляционного анализа состоит в количественном определении тесноты связи между двумя признаками (при парной связи) и между результативным и несколькими факторными признаками при многофакторной связи и статистической оценке надежности установленной связи.

Регрессионный анализ заключается в определении аналитического выражения связи (формы связи), т. е. в выборе математического уравнения, выражающего зависимость между признаками.

Под *формой корреляционной связи* понимают тип аналитической формулы, выражающей зависимость между изучаемыми факторами (признаками). При выборе формы корреляционной связи исходят из экономической сущности изучаемых явлений, простоты аналитической функции, положенной в основание связи, и требований об ограниченном числе параметров.

Уравнение регрессии представляет собой математическую модель, в которой усредненное значение результативного признака y рассматривается как функция одного или нескольких факторных признаков. В первом случае речь идет об уравнении регрессии, характеризующем однофакторную (парную) зависимость между переменными, во втором — о многофакторном регрессионном анализе.

С помощью регрессионного анализа можно осуществлять: построение эмпирических графиков (линий) регрессии; поиск уравнений, позволяющих по эмпирическим данным построить теоретическую), т. е. выровненную линию регрессии; вычисление коэффициентов, позволяющих судить, насколько в среднем результирующая величина изменяется при соответствующих изменениях факторного признака.

Рассмотрим однофакторную линейную регрессию.

Эмпирические графики, отражающие взаимосвязь двух признаков, изображаются в виде *корреляционного поля*. В декартовой системе координат по оси абсцисс откладывают значения факторного признака x, а по оси ординат – результативного y.

Каждой паре значений (x, y) будет соответствовать конкретная точка на плоскости графика. Графическое изображение эмпирических данных может представлять собой множество точек, которое принято называть *диаграммой рассеяния*.

При построении диаграммы возможны различные случаи.

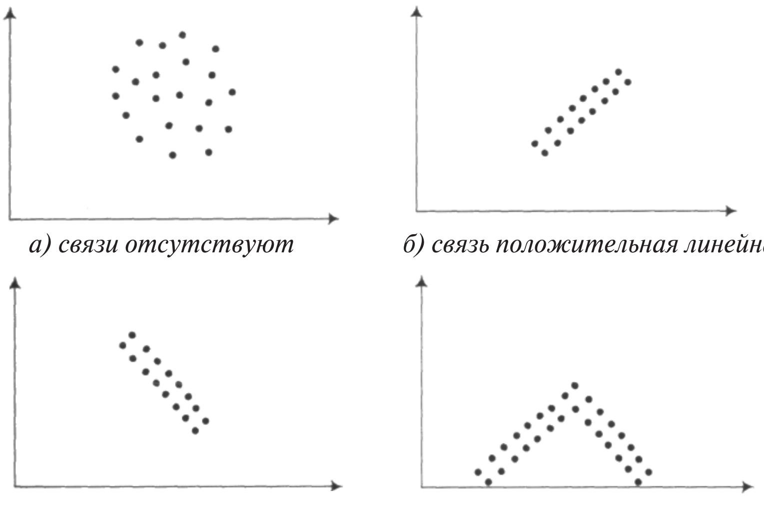
На рис. 2.1 представлено корреляционное поле, состоящее из множества точек, расположенных без какой-либо закономерности, что свидетельствует об отсутствии связи между переменными величинами x и y.

Чем сильнее связь между признаками, тем сильнее будут группироваться эмпирические данные, образуя линию, отражающую конкретную форму связи.

Корреляционное поле, изображенное на рис. 2.1 (б), говорит о наличии положительной линейной зависимости между переменными величинами.

Точки, расположенные на рис. 2.1 (в), свидетельствуют о наличии отрицательной линейной связи.

Расположение точек на рис. 2.1 (г) показывает наличие нелинейной (параболической) зависимости между переменными величинами.



а) связи отсутствуют *б) связь положительная линейная*

в) связь отрицательная линейная *г) связь нелинейная*

Рис. 2.1.

2. Парная регрессия

Парная регрессия представляет собой регрессию между двумя переменными — y и x, т. е. модель вида: $y = \hat{f}(x)$, где y — зависимая переменная (результативный признак); x — независимая, или объясняющая, переменная (признак-фактор). Знак «^» означает, что между переменными x и y нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина y складывается из двух слагаемых: $y = \hat{y}_x + \varepsilon$, где y — фактическое значение результативного признака; $\hat{y}_x$ — теоретическое значение результативного признака, найденное исходя из уравнения регрессии; ε — случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина ε называется также возмущением и включает в себя влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

В парной регрессии выбор вида математической функции $\hat{y}_x = f(x)$ может быть осуществлен тремя методами: графическим; аналитическим, т.е. исходя из теории изучаемой взаимосвязи; экспериментальным.

При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии является наиболее наглядным.

Значительный интерес представляет аналитический метод выбора типа уравнения регрессии. Он основан на изучении материальной природы взаимосвязи исследуемых признаков.

При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется экспериментальным методом путем сравнения величины остаточной дисперсии $\sigma^2_{ост}$, рассчитанной для разных моделей.

Если при этом уравнение регрессии проходит через все точки корреляционного поля, что возможно только при функциональной связи, когда все точки лежат на линии регрессии $\hat{y}_x = f(x)$, то фактические значения результативного признака совпадают с теоретическими $y = \hat{y}_x$, т.е. они полностью обусловлены влиянием фактора x. В этом случае остаточная дисперсия $\sigma_{ост} = 0$.

На практике, как правило, имеет место некоторое рассеяние точек относительно линии регрессии, обусловленое влиянием прочих, не учитываемых в уравнении регрессии факторов. Иными словами, имеют место отклонения фактических данных от теоретических $\left(y-\hat{y}_x\right)$. Величина этих отклонений и лежит в основе расчета остаточной дисперсии:

$\sigma^2_{ост} = \frac{1}{n}\sum\left(y-\hat{y}_x\right)^2$, причем чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к исходным данным.

Для построения уравнения регрессии принять, чтобы число наблюдений должно в 7–8 раз превышать число рассчитываемых параметров при переменной *x*. Это означает, что искать линейную регрессию, имея менее 7 наблюдений, вообще не имеет смысла. Если вид функции усложняется, то требуется увеличение объема наблюдений, так как каждый параметр при *x* должен рассчитываться хотя бы по 7 наблюдениям.

3. Метод наименьших квадратов для построения линейной модели

Рассмотрим простейшую модель парной линейной регрессии. Парная линейная регрессия находит широкое применение в эконометрике благодаря четкой экономической интерпретации ее параметров.

Парная линейная регрессия сводится к нахождению уравнения вида

$$\hat{y}_x = a + b \cdot x \text{ или } y = a + b \cdot x + \varepsilon . \qquad (2.1)$$

Уравнение вида $\hat{y}_x = a + b \cdot x$ позволяет по заданным значениям фактора *x* находить теоретические значения результативного признака, подставляя в него фактические значения фактора *x*.

Построение парной линейной регрессии сводится к оценке ее параметров – *a* и *b*.

Классический подход к оцениванию параметров линейной регрессии основан на применении метода наименьших квадратов (МНК).

Согласно основной идее МНК получают такие оценки параметров *a* и *b*, при которых сумма квадратов отклонений фактических значений результативного признака *y* от теоретических $\hat{y}_x$ минимальна:

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2 = \sum_{i=1}^{n}\varepsilon_i^2 \;\rightarrow\; \min. \tag{2.2}$$

Таким образом, из всего множества линий линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между экспериментальными точками и этой линией была бы минимальной (рис. 2.2).

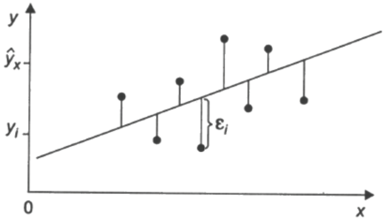


Рис. 2.2.

Из курса математического анализа известно: чтобы найти минимум функции (2.2), надо вычислить частные производные по каждому из параметров *a* и *b* и приравнять их к нулю.

Обозначим $\sum_i \varepsilon_i^2$ через *S (a, b)*, тогда: $S(a, b) = \sum (y - a - b \cdot x)^2$.
Соответственно

$$\begin{cases} \dfrac{dS}{da} = -2\sum y + 2na + 2b\sum x = 0, \\ \dfrac{dS}{db} = -2\sum yx + 2a\sum x + 2b\sum x^2 = 0. \end{cases} \tag{2.3}$$

После несложных преобразований получим следующую систему линейных уравнений для оценки параметров *a* и *b*:

$$\begin{cases} a \cdot n \quad + b \cdot \sum x = \sum y, \\ a \cdot \sum x + b \cdot \sum x^2 = \sum x \cdot y. \end{cases} \tag{2.4}$$

Решая систему уравнений (2.4) методом определителей, после найдем искомые оценки параметров *a* и *b*.

Для нахождения параметров *a* и *b* можно воспользоваться и следующими готовыми формулами, которые следуют непосредственно из решения системы (2.4):

$$a = \overline{y} - b \cdot \overline{x}\,, \qquad b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)}, \tag{2.5}$$

где $\operatorname{cov}(x, y) = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ – ковариация признаков *x* и *y*.

Величину cov *(x, y)* называют *ковариацией. Ковариация* – это числовая характеристика совместного распределения двух случайных величин, равная математическому ожиданию произведения отклонений этих случайных величин от их математических ожиданий.

Величина $\sigma_x^2 = \overline{x^2} - \overline{x}^2$ *– дисперсия признака x.* Дисперсия – характеристика случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины от ее математического ожидания.

Соответственно:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{y_1 + y_2 + \ldots + y_n}{n},$$

$$\overline{y \cdot x} = \frac{1}{n}\sum y \cdot x, \quad \overline{x^2} = \frac{1}{n}\sum x^2,$$

$$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{\text{ï}} (x_i - \overline{x})^2, \qquad \operatorname{var}(y) = \frac{1}{n}\sum_{i=1}^{\text{ï}} (y_i - \overline{y})^2,$$

$$\operatorname{var}(\mathrm{x}) = \sigma_x^2, \qquad \operatorname{var}(\mathrm{y}) = \sigma_y^2,$$

$$\sigma_y = \sqrt{\frac{\sum_{k=1}^{n} (y_k - \overline{y})^2}{n}}, \qquad \sigma_\mathrm{x} = \sqrt{\frac{\sum_{k=1}^{n} (x_k - \overline{x})^2}{n}}, \qquad b \qquad \mathrm{b} = \frac{\overline{\mathrm{yx}} - \overline{\mathrm{y}} \cdot \overline{\mathrm{x}}}{\sigma_\mathrm{x}^2}.$$

4. Линейный коэффициент корреляции

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании парной линейной регрессии в качестве такого показателя выступает линейный *коэффициент корреляции* r_{xy}, который можно рассчитать по следующим формулам:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x,\, y)}{\sigma_x \cdot \sigma_y} \,.. \tag{2.6}$$

Линейный коэффициент корреляции находится в пределах: $-1 \leqslant r_{xy} \leqslant 1$. Чем ближе абсолютное значение r_{xy} к единице, тем сильнее линейная связь между факторами (при $r_{xy} = \pm 1$ имеем строгую функциональную зависимость).

Теснота линейной связи между переменными может быть оценена на основании шкалы Чеддока:

Теснота связи	Значение коэффициента корреляции при наличии:	
	прямой связи	обратной связи
Слабая	0,1–0,3	(–0,3)–(–0,1)
Умеренная	0,3–0,5	(–0,5)–(–0,3)
Заметная	0,5–0,7	(–0,7)–(–0,5)
Высокая	0,7–0,9	(–0,9)–(–0,7)
Весьма высокая	0,9–1	(–1)–(–0,9)

Положительное значение коэффициента корреляции говорит о положительной связи между *x* и *y*, когда с ростом одной из переменных другая тоже растет. Отрицательное значение коэффициента корреляции означает, что с ростом одной из переменных другая убывает, и наоборот: при убывании одной из переменной другая растет.

5. Коэффициент детерминации

Для оценки качества подбора линейной функции, описывающей парную линейную регрессию, рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2, называемый *коэффициентом детерминации*. Коэффициент детерминации характеризует долю дисперсии результативного признака *y*, объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{xy}^2 = 1 - \frac{\sigma_{ост}^2}{\sigma_y^2}, \qquad (2.7)$$

где $\sigma_{ост}^2 = \frac{1}{n}\sum\left(y - \hat{y}_x\right)^2, \quad \sigma_y^2 = \frac{1}{n}\sum\left(y - \bar{y}\right)^2 = \overline{y^2} - \bar{y}^2$.

Соответственно величина $1 - r_{xy}^2$ характеризует долю дисперсии *y*, вызванную влиянием остальных, не учтенных в модели факторов.

6. Проверка значимости и построения доверительных интервалов коэффициентов регрессии

После того как найдено уравнение парной линейной регрессии, проводится оценка значимости как найденного уравнения в целом, так и отдельных его параметров.

Проверить значимость уравнения регрессии – это значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (в рассматриваемом случае одной) для описания зависимой переменной.

Оценка значимости найденного уравнения парной линейной регрессии в целом производится на основе *F*-критерия Фишера. Как правило, проведению *F*-критерия Фишера предшествует дисперсионный анализ.

Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной y от среднего значения $\overline{y}$ раскладывается на две части – «объясненную» (или «факторную») и «необъясненную»:

$$\sum(y-\overline{y})^2=\sum\left(\hat{y}_x-\overline{y}\right)^2+\sum\left(y-\hat{y}_x\right)^2, \tag{2.8}$$

где $\sum(y-\overline{y})^2$ – общая сумма квадратов отклонений; $\sum\left(\hat{y}_x-\overline{y}\right)^2$ – сумма квадратов отклонений, объясненная регрессией; $\sum\left(y-\hat{y}_x\right)^2$ – остаточная сумма квадратов отклонений, характеризующая влияние не учтенных в модели факторов.

Результаты дисперсионного анализа представляют в виде таблицы 2.1 (n – число наблюдений, m – число параметров при переменной x).

В расчетах используются не сами дисперсии, а дисперсии, рассчитанные на одну степень свободы. Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину *F*-критерия Фишера:

$$\mathrm{F}=\frac{\mathrm{S}^2_{\text{факт}}}{\mathrm{S}^2_{\text{ост}}}\ . \tag{2.9}$$

Таблица 2.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum\left(y-\bar{y}\right)^2$	n	$S^2_{\text{общ}} = \dfrac{\sum(y-\bar{y})^2}{n-1}$
Факторная	$\sum\left(\hat{y}_x-\bar{y}\right)$	m	$S^2_{\text{фактич}} = \dfrac{\sum(\hat{y}_x-\bar{y})^2}{m}$
Остаточная	$\sum\left(y-\hat{y}_x\right)^2$	$n-m-1$	$S^2_{\text{ост}} = \dfrac{\sum(y-\hat{y}_x)^2}{n-m-1}$

Выдвигается гипотеза H_0 о статистической незначимости найденного уравнения регрессии.

Если нулевая гипотеза H_0 справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Если H_0 несправедлива, то факторная дисперсия превышает остаточную в несколько раз.

Английским статистиком Снедекором разработаны таблицы критических значений F-отношений при разных уровнях значимости нулевой гипотезы и различном числе степеней свободы.

С учетом таблиц F-отношений Снедекора фактическое значение F-критерия Фишера (2.9) сравнивается с табличным значением $F_{\text{табль.}}$ $(\alpha; k_1; k_2)$; $k_1 = m$, $k_2 = n-m-1$; $\alpha = 1-p$; p – доверительная вероятность; α – уровень значимости.

Если фактическое значение F-критерия больше табличного, то признается статистическая значимость найденного уравнения в целом. В этом случае нулевая гипотеза об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи:

$F > F_{табл}$, H_0 отклоняется.

Если же величина F окажется меньше табличной, то вероятность нулевой гипотезы выше заданного уровня значимости α и она не может быть отклонена без риска сделать неправильный вывод о наличии связи. В этом случае уравнение регрессии считается статистически незначимым:

$F < F_{табл}$, H_0 не отклоняется.

Для парной линейной регрессии $m = 1$, поэтому

$$F = \frac{S_{\text{факт}}}{S^2_{\text{ост}}} = \frac{\sum\left(\hat{y}_{x} - \bar{y}\right)^2}{\sum\left(y - \hat{y}\right)^2} \cdot \left(n-2\right). \qquad (2.10)$$

Величина F-критерия связана с коэффициентом детерминации r^2_{xy}, и ее можно рассчитать по следующей формуле:

$$F = \frac{r^2_{xy}}{1 - r^2_{xy}} \cdot \left(n-2\right). \qquad (2.11)$$

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка: m_b и m_a.

Стандартная ошибка коэффициента регрессии определяется по формуле:

$$m_b = \sqrt{\frac{S^2_{\text{ост}}}{\sum\left(x - \bar{x}\right)^2}} = \frac{S_{\text{ост}}}{\sigma_{x} \cdot \sqrt{n}}, \qquad (2.12)$$

где $S^2_{ост} = \dfrac{\sum\left(y - \hat{y}_x\right)^2}{n-2}$ — остаточная дисперсия на одну степень свободы.

Величина стандартной ошибки совместно с t-распределением Стьюдента при $n - 2$ степенях свободы применяется для проверки существенности гипотезы H_0 о значимости коэффициента регрессии и для расчета его доверительного интервала.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется фактическое значение t-критерия Стьюдента: $t_b = \dfrac{b}{m_b}$, которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы ($n - 2$).

Отношение коэффициента регрессии к его стандартной ошибке дает t-статистику, которая подчиняется статистике Стьюдента при (n–2) степенях свободы. Эта статистика применяется для проверки статистической значимости коэффициента регрессии и для расчета его доверительных интервалов.

Если фактическое значение t-критерия превышает табличное, гипотезу H_0 о несущественности коэффициента регрессии можно отклонить. Доверительный интервал для коэффициента регрессии определяется как $b \pm t \cdot m_b$.

Границы доверительного интервала для коэффициента регрессии не должны содержать противоречивых результатов, например, $-1{,}5 \leq b \leq 0{,}8$. Такого рода запись указывает, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть, и коэффициент признается незначимым.

Стандартная ошибка параметра a определяется по формуле:

$$m_a = \sqrt{S^2_{ост} \cdot \frac{\sum x^2}{n\sum\left(x-\bar{x}\right)^2}} = S_{ост} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n}. \tag{2.13}$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии. Вычисляется t-критерий: $t_a = \dfrac{a}{m_a}$, его величина сравнивается с табличным значением при $n - 2$ степенях свободы.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции m_r:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}. \tag{2.14}$$

Фактическое значение t-критерия Стьюдента определяется как $t_r = \dfrac{r}{m_r}$.

Существует связь между t-критерием Стьюдента и F-критерием Фишера:

$$t_b = t_r = \sqrt{F}. \tag{2.15}$$

В прогнозных расчетах по уравнению парной линейной регрессии определяется предсказываемое $\hat{y}_p$ значение как точечный прогноз $\hat{y}_x$ при $x_p = x_k$, т.е. путем подстановки в уравнение регрессии $\hat{y}_x = a + b \cdot x$ соответствующего значения x. Однако точечный прогноз явно не реален. Поэтому он дополняется расчетом стандартной ошибки $\hat{y}_p$, т.е. $m_{\hat{y}_p}$, и, соответственно, интервальной оценкой прогнозного значения $\hat{y}_p$:

$$\hat{y}_p - \Delta_{\hat{y}_p} \leq \hat{y}_p \leq \hat{y}_p + \Delta_{\hat{y}_p},$$

где $\Delta_{\hat{y}_p} = m_{\hat{y}_p} \cdot t_{табл}$, а $m_{\hat{y}_p}$ – средняя ошибка прогнозируемого индивидуального значения:

$$m_{\hat{y}_p} = S_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2}}.$$

7. Интерпретация уравнений

Параметр b называется *коэффициентом регрессии*. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Формально a – значение y при $x = 0$. Если признак-фактор x не может иметь нулевого значения, то вышеуказанная трактовка свободного члена a не имеет смысла, т.е. параметр a может не иметь экономического содержания.

8. Средняя ошибка аппроксимации

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100\%. \qquad (2.16)$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / Под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

3. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А.А. Пересецкий. – Москва: Дело, 1997. – 442 с.

4. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 324 с.

5. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

План:

1. Множественная регрессия. Отбор факторов при построении уравнения множественной регрессии.
2. Оценка параметров уравнения множественной регрессии.
3. Множественный коэффициент корреляции. Его значимость.
4. Частные коэффициенты корреляции. Их выборочные оценки.
5. Проверка значимости уравнения множественной регрессии в целом.
6. Частный F-критерий.

1. Множественная регрессия. Отбор факторов при построении уравнения множественной регрессии

При моделировании экономических процессов парная регрессия дает хороший результат, если влиянием других факторов кроме одного, воздействующих на объект исследования, можно пренебречь. В том случае, когда этим влиянием пренебречь нельзя, следует попытаться выявить влияние других факторов, введя их в модель, т. е. построить уравнение множественной регрессии

$$y = \hat{f}\left(x_1,\ x_2,\ \ldots,\ x_m\right),$$

где y – зависимая переменная (результативный признак), x_i – независимые, или объясняющие, переменные (признаки-факторы).

Возможны как линейные, так и нелинейные виды уравнений множественной регрессии:

В силу четкой интерпретации параметров наиболее широко используется линейная модель. В линейной множественной регрессии $\hat{y}_x = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m$ параметры при x называются *коэффициентами «чистой» регрессии*, они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне.

Построение уравнения множественной регрессии начинается с решения вопроса *о спецификации модели*. Он включает в себя два круга вопросов: отбор факторов и выбор вида уравнения регрессии.

Включение в уравнение множественной регрессии того или иного набора факторов связано прежде всего с представлением исследователя о природе взаимосвязи моделируемого показателя с другими экономическими явлениями. Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

1. Они должны быть количественно измеримы. Если же необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.
2. Факторы не должны быть интеркоррелированы и тем более находиться в точной функциональной связи.

Отбор факторов прежде всего производится на основе качественного теоретико-экономического анализа. Однако теоретический анализ часто не позволяет однозначно ответить на вопрос о количественной взаимосвязи рассматриваемых признаков и целесообразности включения того или иного фактора в модель. Поэтому отбор факторов обычно осуществляется в две стадии: на первой подбираются факторы исходя из сущности проблемы; на второй – на основе матрицы показателей корреляции определяют статистики для параметров регрессии.

Коэффициенты *интеркорреляции* (т. е. корреляции между объясняющими переменными) позволяют исключать из модели дублирующие факторы. Считается, что две переменные явно *коллинеарны*, т. е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0{,}7$.

Если факторы явно *коллинеарны*, то есть находятся между собой в линейной зависимости, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. В этом случае предпочтение отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами. В этом требовании проявляется специфика множественной регрессии как метода исследования комплексного воздействия факторов в условиях их независимости друг от друга.

По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии *мультиколлинеарности факторов*, когда более чем два фактора связаны между собой линейной зависимостью, т. е. имеет место совокупное воздействие факторов друг на друга.

Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов корреляции между факторами. Чем ближе к нулю определитель этой матрицы, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И наоборот: чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Существует ряд подходов преодоления сильной межфакторной корреляции. Самый простой путь устранения мультиколлинеарности состоит в исключении из модели одного или нескольких факторов. Другой подход связан с преобразованием факторов, при котором уменьшается корреляция между ними.

При отборе факторов также рекомендуется пользоваться следующим правилом: число включаемых факторов обычно в 6–7 раз меньше объема совокупности, по которой строится регрессия. Если это соотношение нарушено, то число степеней свободы остаточной дисперсии очень мало. Это приводит к тому, что параметры уравнения регрессии оказываются статистически не значимыми, а F-критерий меньше табличного значения.

2. Оценка параметров уравнения множественной регрессии

Рассмотрим линейную модель множественной регрессии

$$\hat{y} = a + b_1x_1 + b_2x_2 + ... + b_mx_m + \varepsilon. \qquad (3.1)$$

Классический подход к оцениванию параметров линейной модели множественной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных $\hat{y}$ минимальна:

$$\sum_i \left(y_i - \hat{y}_{x_i} \right)^2 \quad \to \quad \min. \qquad (3.2)$$

Как известно из курса математического анализа, для того чтобы найти экстремум функции нескольких переменных, надо вычислить частные производные первого порядка по каждому из параметров и приравнять их к нулю.

Имеем функцию $m + 1$ аргумента:

$$S(a,\, b_1,\, b_2,\, ...,\, b_m) = \sum \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m \right)^2 .$$

Находим частные производные первого порядка и после элементарных преобразований приходим к *системе линейных нормальных уравнений* для нахождения параметров линейного уравнения множественной регрессии (3.1):

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 + ... + b_m\sum x_m = \sum y, \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 + ... + b_m\sum x_1x_m = \sum yx_1, \\ .... \\ a\sum x_m + b_1\sum x_1x_m + b_2\sum x_2x_m + ... + b_m\sum x_m^2 = \sum yx_m. \end{cases} \quad (3.3)$$

Решение этой системы уравнений осуществляется методом определителей:

$$a = \frac{\Delta a}{\Delta}, \quad b_1 = \frac{\Delta b_1}{\Delta}, \quad .... \; b_m = \frac{\Delta b_m}{\Delta},$$

где Δ – определитель системы, Δa, Δb_1 .... Δb_m – частные определители.

Для двухфакторной модели данная система будет иметь вид:

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 = \sum y, \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 = \sum yx_1, \\ a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 = \sum yx_2. \end{cases}$$

Метод наименьших квадратов применим и к уравнению множественной регрессии, записанному в *стандартизированном масштабе*:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + ... + \beta_m t_{x_m} + \varepsilon, \quad (3.4)$$

где t_y, t_{x_1}, ..., t_{x_m} – стандартизированные переменные:

$t_y = \dfrac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \dfrac{x_i - \bar{x}_i}{\sigma_{x_i}}$, для которых среднее значение равно нулю: $\bar{t}_y = \bar{t}_{x_i} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$; β_i – стандартизированные коэффициенты регрессии.

Стандартизованные коэффициенты регрессии показывают, на сколько сигм изменится в среднем результат, если соответствующий фактор x_i изменится на одну единицу при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. *В этом основное достоинство стандартизованных коэффициентов регрессии, в отличие от коэффициентов «чистой» регрессии, которые не сравнимы между собой.*

Применяя МНК к уравнению множественной регрессии в стандартизированном масштабе, получим систему нормальных уравнений вида:

$$\begin{cases} r_{yx_1} = \beta_1 \qquad + \beta_2 r_{x_1x_2} + \beta_3 r_{x_1x_3} + \ldots + \beta_m r_{x_1x_m}, \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2 \qquad + \beta_3 r_{x_1x_3} + \ldots + \beta_m r_{x_1x_m}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{yx_m} = \beta_1 r_{x_1x_m} + \beta_2 r_{x_2x_m} + \beta_3 r_{x_3x_m} + \ldots + \beta_m, \end{cases} \tag{3.5}$$

где r_{yx_i} и $r_{x_ix_j}$ – коэффициенты парной и межфакторной корреляции.

Решая ее методом определителей, находят параметры – стандартизованные коэффициенты регрессии или β-*коэффициенты*.

Коэффициенты «чистой» регрессии b_i связаны со стандартизованными коэффициентами регрессии β_i следующим образом:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \tag{3.6}$$

Поэтому можно переходить от уравнения регрессии в стандартизованном масштабе (3.4) к уравнению регрессии в естественной форме (натуральном масштабе переменных) (3.1), при этом параметр a определяется как

$$a = \overline{y} - b_1\overline{x}_1 - b_2\overline{x}_2 - \ldots - b_m\overline{x}_m.$$

Рассмотренный смысл стандартизованных коэффициентов регрессии позволяет их использовать при отсеве факторов – из модели исключаются факторы с наименьшим значением β_i.

На основе линейного уравнения множественной регрессии

$$\hat{y} = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + \varepsilon \tag{3.7}$$

могут быть найдены *частные уравнения* регрессии:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_m} = \hat{f}(x_1), \\ y_{x_2 \cdot x_1, x_3, \ldots, x_m} = \hat{f}(x_2), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_m \cdot x_1, x_2, \ldots, x_{m-1}} = \hat{f}(x_m), \end{cases} \tag{3.8}$$

т. е. уравнения регрессии, которые связывают результативный признак с соответствующим фактором x_i при закреплении остальных факторов на среднем уровне. В развернутом виде систему (3.8) можно переписать в виде:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_m} = a + b_1 x_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m + \varepsilon, \\ y_{x_2 \cdot x_1, x_3, \ldots, x_m} = a + b_1 \bar{x}_1 + b_2 x_2 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m + \varepsilon, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_m \cdot x_1, x_2, \ldots, x_{m-1}} = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_m x_m + \varepsilon. \end{cases}$$

При подстановке в эти уравнения средних значений соответствующих факторов они принимают вид парных уравнений линейной регрессии, т. е. имеем:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_m} = A_1 + b_1 x_1, \\ y_{x_2 \cdot x_1, x_3, \ldots, x_m} = A_2 + b_2 x_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_m \cdot x_1, x_2, \ldots, x_{m-1}} = A_m + b_m x_m, \end{cases} \tag{3.9}$$

где $\begin{cases} A_1 = a + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m, \\ A_2 = a + b_1 \bar{x}_1 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ A_m = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_{m-1} x_{m-1}. \end{cases}$

В отличие от парной регрессии частные уравнения регрессии характеризуют изолированное влияние фактора на результат, так как другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии определять *частные коэффициенты эластичности*:

$$\acute{Y}_{y_{x_i}} = b_i \cdot \frac{x_i}{\hat{y}_{x_i \cdot x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m}}, \qquad (3.10)$$

где b_i – коэффициент регрессии для фактора x_i в уравнении множественной регрессии, $\hat{y}_{x_i \cdot x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots, x_m}$ – частное уравнение регрессии.

Наряду с частными коэффициентами эластичности могут быть найдены *средние по совокупности показатели эластичности*:

$$\bar{\acute{Y}}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}, \qquad (3.11)$$

которые показывают, на сколько процентов в среднем изменится результат при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и, соответственно, так же ранжировать факторы по силе их воздействия на результат.

3. Множественный коэффициент корреляции. Его значимость

Рассмотрим методы проверки существенности факторов и показатели качества множественной регрессии.

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – показателя детерминации.

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, или, иначе, оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции:

$$R_{yx_1x_2 \ldots x_m} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}}, \qquad (3.12)$$

где $\sigma^2_{ост}$ – остаточная дисперсия для уравнения $y = f(x_1, x_2 \ldots x_m)$, σ^2_y – общая дисперсия результативного признака.

Границы изменения индекса множественной корреляции – от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина индекса множественной корреляции должна быть больше или равна максимальному парному индексу корреляции:

$$R_{yx_1x_2\ldots x_m} \geq r_{yx_i\,(\max)} \quad \left(i=\overline{1,m}\right).$$

Расчет индекса множественной корреляции предполагает определение уравнения множественной регрессии и на его основе остаточной дисперсии:

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum\left(y-\hat{y}_{x_1x_2\ldots\ldots x_m}\right)^2. \tag{3.13}$$

Можно пользоваться следующей формулой для расчета индекса множественной детерминации:

$$R^2_{yx_1x_2\ldots x_m} = 1-\frac{\sum\left(y-\hat{y}_{x_1x_2\ldots x_m}\right)^2}{\sum\left(y-\overline{y}\right)^2}. \tag{3.14}$$

При линейной зависимости признаков формула индекса множественной корреляции может быть представлена в следующем виде:

$$R_{yx_1x_2\ldots x_m} = \sqrt{\sum\beta_{x_i}\cdot r_{yx_i}}\,, \tag{3.15}$$

где β_{x_i} – стандартизованные коэффициенты регрессии, r_{yx_i} – парные коэффициенты корреляции результата с каждым фактором.

Формула индекса множественной корреляции для линейной регрессии получила название *линейного коэффициента множественной корреляции*, или, что то же самое, *совокупного коэффициента корреляции.*

Для уравнения множественной $y = a + b_1x_1 + b_2x_2 + \ldots\ldots b_px_p + \varepsilon$ можно определить совокупный коэффициент корреляции через матрицу парных коэффициентов корреляции:

$$R_{yx_1x_2,\ldots,x_p} = \sqrt{1-\frac{\Delta r}{\Delta r_{11}}}\,, \tag{3.16}$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \ldots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \ldots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \ldots & r_{x_2x_p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \ldots & 1 \end{vmatrix}$$

– определитель матрицы парных коэффициентов корреляции,

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$

– определитель матрицы межфакторной корреляции.

В рассмотренных показателях множественной корреляции (индекс и коэффициент) используется остаточная дисперсия, которая имеет систематическую ошибку в сторону преуменьшения, тем более значительную, чем больше параметров определяется в уравнении регрессии при заданном объеме наблюдений n. Если число параметров при x_i равно m и приближается к объему наблюдений, то остаточная дисперсия будет близка к нулю и коэффициент (индекс) корреляции приблизится к единице даже при слабой связи факторов с результатом. Для того чтобы не допустить возможного преувеличения тесноты связи, используется скорректированный индекс (коэффициент) множественной корреляции.

Скорректированный индекс множественной корреляции содержит поправку на число степеней свободы, а именно: остаточная сумма квадратов $\sum\left(y-\hat{y}_{x_1x_2\dots x_m}\right)^2$ делится на число степеней свободы остаточной вариации $(n - m - 1)$, а общая сумма квадратов отклонений $\sum(y-\bar{y})^2$ – на число степеней свободы в целом по совокупности $(n - 1)$.

Формула для расчета скорректированного индекса множественной детерминации имеет вид:

$$\hat{R}^2 = 1 - \frac{\sum\left(y-\hat{y}\right)^2 \big/ (n-m-1)}{\sum(y-\bar{y}) / (n-1)}, \tag{3.17}$$

где m – число параметров при переменных x, n – число наблюдений.

Поскольку $\dfrac{\sum\left(y-\hat{y}_{x_1x_2\dots x_m}\right)^2}{\sum(y-\bar{y})^2} = 1 - R^2$, то величину скорректированного индекса детерминации можно представить в виде:

$$\hat{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-m-1} \cdot \qquad (3.17a)$$

Чем больше величина m, тем сильнее различия $\hat{R}^2$ и R^2.

Как было показано выше, ранжирование факторов, участвующих во множественной линейной регрессии, может быть проведено через стандартизованные коэффициенты регрессии (β – коэффициенты). Эта же цель может быть достигнута с помощью частных коэффициентов корреляции (для линейных связей). Кроме того, частные коэффициенты корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель можно доказать, опираясь на величину показателей частной корреляции, которые будут рассмотрены ниже.

4. Частные коэффициенты корреляции. Их выборочные оценки

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель.

В общем виде при наличии m – факторов для уравнения

$$y = a + b_1 x_1 + b_2 x_2 ... + b_m \bar{x}_m + \varepsilon$$

коэффициент частной корреляции, измеряющий влияние на y фактора x_i, при неизменном уровне других факторов, можно определить по формуле:

$$r_{yx_i \cdot x_1 x_2 ... x_{i-1} x_{i+1} ... x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 ... x_i ... x_m}}{1 - R^2_{yx_1 x_2 ... x_{i-1} x_{i+1} ... x_m}}}, \qquad (3.18)$$

где $R^2_{yx_1 x_2 ... x_i ... x_m}$ – множественный коэффициент детерминации всех m факторов с результатом; $R^2_{yx_1 x_2 ... x_{i-1} x_{i+1} ... x_m}$ – тот же показатель детерминации, но без введения в модель фактора x_i.

При двух факторах формула (3.18) примет вид:

$$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}; \qquad r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}}. \tag{3.18a}$$

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, $r_{yx_1 \cdot x_2}$ – коэффициент частной корреляции первого порядка. Соответственно коэффициенты парной корреляции называются коэффициентами нулевого порядка. Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле:

$$r_{yx_i \cdot x_1x_2 \ldots x_{i-1}x_{i+1} \ldots x_m} = \frac{r_{yx_i \cdot x_1x_2 \ldots x_{i-1}x_{i+1} \ldots x_{m-1}} - r_{yx_m \cdot x_1x_2 \ldots x_{m-1}} \cdot r_{x_ix_m \cdot x_1x_2 \ldots x_{i-1}x_{i+1} \ldots x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_m \cdot x_1x_2 \ldots x_{m-1}}\right) \cdot \left(1 - r^2_{x_ix_m \cdot x_1x_2 \ldots x_{i-1}x_{i+1} \ldots x_{m-1}}\right)}}. \tag{3.19}$$

При двух факторах эта формула примет вид:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}}, \quad r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}}. \tag{3.19a}$$

Для уравнения регрессии с тремя факторами частные коэффициенты корреляции второго порядка определяются на основе частных коэффициентов корреляции первого порядка. Так, по уравнению $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$ возможно исчисление трех частных коэффициентов корреляции второго порядка:

$r_{yx_1 \cdot x_2x_3}$, $r_{yx_2 \cdot x_1x_3}$, $r_{yx_3 \cdot x_1x_2}$,

каждый из которых определяется по рекуррентной формуле. Например, при $i = 1$ имеем формулу для расчета $r_{yx_1 \cdot x_2x_3}$:

$$r_{yx_1 \cdot x_2x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2} \cdot r_{x_1x_3 \cdot x_2}}{\sqrt{\left(1 - r^2_{yx_3 \cdot x_2}\right) \cdot \left(1 - r^2_{x_1x_3 \cdot x_2}\right)}}. \tag{3.20}$$

Рассчитанные по рекуррентной формуле частные коэффициенты корреляции изменяются в пределах от –1 до +1, а по формулам через множественные коэффициенты детерминации – от 0 до 1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом.

Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде. Если из стандартизованного уравнения регрессии $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \beta_3 t_{x_3} + \varepsilon$ следует, что $\beta_1 > \beta_2 > \beta_3$, т. е. по силе влияния на результат порядок факторов таков: x_1, x_2, x_3, то этот же порядок факторов определяется и по соотношению частных коэффициентов корреляции, $r_{yx_1 \cdot x_2 x_3} > r_{yx_2 \cdot x_1 x_3} > r_{yx_3 \cdot x_1 x_2}$.

Согласованность частной корреляции и стандартизованных коэффициентов регрессии наиболее отчетливо видна из сопоставления формул для их расчета при двухфакторном анализе. Для уравнения регрессии в стандартизованном масштабе

$$\hat{t}_y = \beta_{x_1} \cdot t_{x_1} + \beta_{x_2} \cdot t_{x_2}$$

β-коэффициенты могут быть определены по формулам, полученным из решения системы нормальных уравнений:

$$\beta_{x_1} = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r^2_{x_1 x_2}}, \qquad \beta_{x_2} = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r^2_{x_1 x_2}}.$$

Сравнивая их с рекуррентными формулами расчета частных коэффициентов корреляции $r_{yx_1 \cdot x_2}$, $r_{yx_2 \cdot x_1}$, можно видеть, что

$$r_{yx_1 \cdot x_2} = \beta_{x_1} \cdot \sqrt{\frac{1 - r^2_{x_1 x_2}}{1 - r^2_{yx_2}}}, \quad r_{yx_2 \cdot x_1} = \beta_{x_2} \cdot \sqrt{\frac{1 - r^2_{x_1 x_2}}{1 - r^2_{yx_1}}}.$$

Иными словами, в двухфакторной модели частные коэффициенты корреляции — это стандартизованные коэффициенты регрессии, умноженные на корень квадратный из соотношения долей остаточных дисперсий фиксируемого фактора на фактор и на результат.

Из приведенных выше формул частных коэффициентов корреляции видна связь этих показателей с совокупным коэффициентом корреляции.

Зная частные коэффициенты корреляции (последовательно первого, второго и более высокого порядка), можно определить совокупный коэффициент корреляции по формуле:

$$R_{yx_1 x_2 \ldots x_m} = \sqrt{1 - \left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{yx_2 \cdot x_1}\right) \cdot \left(1 - r^2_{yx_3 \cdot x_1 x_2}\right) \cdot \ldots \cdot \left(1 - r^2_{yx_m \cdot x_1 x_2 \ldots x_{m-1}}\right)}. \quad (3.21)$$

В частности, для двухфакторного уравнения формула (3.21) принимает вид:

$$R_{yx_1x_2} = \sqrt{1 - \left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{yx_2 \cdot x_1}\right)}.$$

При полной зависимости результативного признака от исследуемых факторов коэффициент совокупного их влияния равен единице. Из единицы вычитается доля остаточной вариации результативного признака $(1-r^2)$, обусловленная последовательно включенными в анализ факторами. В результате подкоренное выражение характеризует совокупное действие всех исследуемых факторов.

5. Проверка значимости уравнения множественной регрессии в целом

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью F-критерия Фишера:

$$F = \frac{S_{\text{факт}}}{S_{\text{ост}}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}, \tag{3.22}$$

где $S_{факт}$ — факторная сумма квадратов на одну степень свободы; $S_{ост}$ — остаточная сумма квадратов на одну степень свободы; R^2 — коэффициент (индекс) множественной детерминации; m — число параметров при переменных x (в линейной регрессии совпадает с числом включенных в модель факторов); n — число наблюдений.

Оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в регрессионную модель. Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, может существенно увеличивать долю объясненной вариации результативного признака. Кроме того, при наличии в модели нескольких факторов они могут вводиться в модель в разной последовательности. Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель. Мерой для оценки включения фактора в модель служит *частный F-критерий*, т. е. F_{xi}.

6. Частный F-критерий

Частный F-критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом. В общем виде для фактора x_i частный F-критерий определится как

$$F_{x_i} = \frac{R^2_{yx_1 \ldots x_i \ldots x_m} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}}{1 - R^2_{yx_1 \ldots x_i \ldots x_m}} \cdot \frac{n - m - 1}{1}, \tag{3.23}$$

где $R^2_{yx_1 \ldots x_i \ldots x_m}$ — коэффициент множественной детерминации для модели с полным набором факторов; $R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}$ — тот же показатель, но без включения в модель фактора x_i; n — число наблюдений; m — число параметров в модели (без свободного члена).

Фактическое значение частного F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: 1 и $n - m - 1$. Если фактическое значение F_{x_i} превышает $F_{табл} = (\alpha, k_1, k_2)$, то дополнительное включение фактора x_i в модель статистически оправданно и коэффициент чистой регрессии b_i при факторе x_i статистически значим. Если же фактическое значение F_{x_i} меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объясненной вариации признака y, следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически не значим.

Для двухфакторного уравнения частные F-критерии имеют вид:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n - 3), \quad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n - 3). \tag{3.23а}$$

С помощью частного F-критерия можно проверить значимость всех коэффициентов регрессии в предположении, что каждый соответствующий фактор x_i вводился в уравнение множественной регрессии последним.

Частный F-критерий оценивает значимость коэффициентов чистой регрессии. Зная величину F_{x_i}, можно определить и t-критерий для коэффициента регрессии при i-м факторе, t_{b_i}, а именно:

$$t_{b_i} = \sqrt{F_{x_i}}. \tag{3.24}$$

Оценка значимости коэффициентов чистой регрессии по t-критерию Стьюдента может быть проведена и без расчета частных F-критериев. В этом случае, как и в парной регрессии, для каждого фактора используется формула:

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \tag{3.25}$$

где b_i – коэффициент чистой регрессии при факторе x_i, m_{b_i} – средняя квадратическая (стандартная) ошибка коэффициента регрессии b_i.

Для уравнения множественной регрессии $\hat{y} = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m$ средняя квадратическая ошибка коэффициента регрессии может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1 \ldots x_m}}}{\sigma_{x_i} \sqrt{1 - R^2_{x_i x_1 \ldots x_m}}} \cdot \frac{1}{\sqrt{n - m - 1}}, \tag{3.26}$$

где σ_y – среднее квадратическое отклонение для признака y, σ_{x_i} – среднее квадратическое отклонение для признака x_i, $R^2_{yx_1 \ldots x_m}$ – коэффициент детерминации для уравнения множественной регрессии, $R^2_{x_i x_1 \ldots x_m}$ – коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии; $n - m - 1$ – число степеней свободы для остаточной суммы квадратов отклонений.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

3. Айвазян, С. А. Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001. – 300 с.

4. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва: Дело, 1997. – 442 с.

5. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 324 с.

6. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

Лекция 4

ПРЕДПОСЫЛКИ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ. ГЕТЕРОСКЕДАСТИЧНОСТЬ. АВТОКОРРЕЛЯЦИЯ

План:

1. Предпосылки метода наименьших квадратов.
2. Гетероскедастичность. Суть и последствия гетероскедастичности. Тест ранговой корреляции Спирмера.
3. Автокорреляция. Суть и причины автокорреляция. Последствия и обнаружение автокорреляции. Метод устранения автокорреляции.
4. Обобщенная линейная модель множественной регрессии и обобщенный метод наименьших квадратов.

1. Предпосылки метода наименьших квадратов

Предпосылками метода наименьших квадратов являются условия, необходимые для получения несмещенных, состоятельных и эффективных оценок, соблюдение которых необходимо для получения достоверных результатов регрессии. Для этого необходимо выполнение следующих пяти предпосылок:

- случайный характер остатков;
- нулевая средняя величина остатков, не зависящая от независимой переменной;
- гомоскедастичность – дисперсия каждого отклонения одинакова для всех значений независимой переменной;
- отсутствие автокорреляции остатков (значения остатков распределены независимо друг от друга);
- остатки подчиняются нормальному закону распределения.

2. Гетероскедастичность. Суть и последствия гетероскедастичности. Тест ранговой корреляции Спирмера

При практическом решении задач регрессионного анализа с помощью МНК обращают серьезное внимание на проблемы, связанные с выполнимостью свойств случайных отклонений остатков. Учитывается, что свойства оценок коэффициентов регрессии напрямую зависят от свойств случайного члена в уравнении регрессии. Для получения качественных оценок необходимо следить за выполнимостью предпосылок МНК (условий Гаусса–Маркова), так как при их нарушении МНК может давать оценки с плохими статистическими свойствами.

Одной из ключевых предпосылок МНК является условие постоянства дисперсий случайных отклонений. Выполнимость данной предпосылки называется *гомоскедастичностью (постоянством дисперсии отклонений)*. Невыполнимость данной предпосылки называется *гетероскедастичностью (непостоянством дисперсий отклонений)*.

Проанализируем суть гетероскедастичности, ее причины и последствия и способ смягчения этих последствий.

Суть гетероскедастичности

На рис. 4.1 приведены два примера линейной регрессии — зависимости потребления C от дохода I.

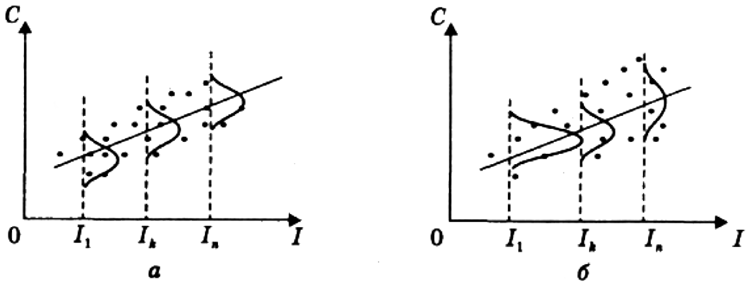


Рис. 4.1

В обоих случаях с ростом дохода растет среднее значение потребления. Но если на рис. 4.1а дисперсия потребления остается одной и той же для различных уровней дохода, то на рис. 4.1б 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 существен для большего уровня дохода. Люди с большим доходом имеют больший простор для его распределения. Реалистичность данной ситуации не вызывает сомнений. Разброс значений потребления вызывает разброс точек наблюдения относительно линии регрессии, что и определяет дисперсию случайных отклонений. Динамика изменения дисперсий (распределений) отклонений для данного примера проиллюстрирована на рис. 4.2. При гомоскедастичности (рис. 4.2а) дисперсии постоянны, а при гетероскедастичности (рис. 4.2б) дисперсии изменяются (в нашем примере увеличиваются).

Проблема гетероскедастичности характерна для перекрестных данных и довольно редко встречается при рассмотрении временных рядов. Это можно объяснить следующим. При перекрестных данных учитываются экономические субъекты (потребители, фирмы, отрасли, страны), имеющие различные доходы, размеры, потребности и т. д. В этом случае возможны проблемы, связанные с эффектом масштаба. Во временных рядах обычно рассматриваются одни и те же показатели в различные моменты времени (например, ВВП, чистый экспорт, темпы инфляции и т. д. в определенном регионе за определенный период времени). Однако при увеличении (уменьшении) рассматриваемых показателей с течением времени может возникнуть проблема гетероскедастичности.

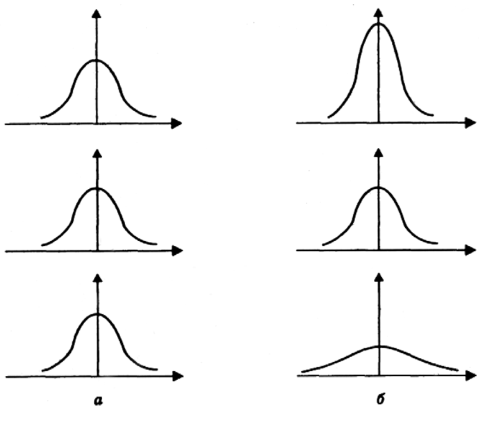


Рис. 4.2

Последствия гетероскедастичности

При рассмотрении классической линейной регрессионной модели МНК дает наилучшие линейные несмещенные оценки лишь при выполнении ряда предпосылок, одной из которых является постоянство дисперсии отклонений (гомоскедастичность для всех наблюдений).

При невыполнимости данной предпосылки (при гетероскедастичности) последствия применения МНК будут следующими:

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

2. Оценки не будут эффективными (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.

4. Вследствие вышесказанного все выводы, получаемые на основе соответствующих F- и t-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 являющихся.

Обнаружение гетероскедастичности

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

Обнаружение гетероскедастичности в каждом конкретном случае является довольно сложной задачей. Какого-либо однозначного метода определения гетероскедастичности не существует.

Рассмотрим наиболее популярный тест ранговой корреляции Спирмена.

Тест ранговой корреляции Спирмена

При использовании данного теста предполагается, что дисперсия отклонения будет либо увеличиваться, либо уменьшаться с увеличением значений x. Поэтому для регрессии, построенной по МНК, абсолютные величины отклонений ε_i и значения x_i будут коррелированы. Значения ε_i и x_i ранжируются (упорядочиваются по величинам). Затем определяется коэффициент ранговой корреляции:

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

где d_i – разность между рангами $\varepsilon_i x_i$, $i=1, 2, 3, 4...... n$; n – число наблюдений.

Доказано, что если коэффициент корреляции r_{xe} для генеральной совокупности равен нулю, то статистика

$$t = \frac{r_{xe}\sqrt{n-2}}{\sqrt{1-r_{xe}^2}}$$

имеет распределение Стьюдента с числом степеней свободы
$v = n - 2$.

Следовательно, если наблюдаемое значение t-статистики, вычисленное по формуле для $r_{x,e}$, превышает определяемое по таблице критических точек распределения Стьюдента, то необходимо отклонить гипотезу о равенстве нулю коэффициента корреляции $r_{x,e}$, α следовательно, и об отсутствии гетероскедастичности. В противном случае гипотеза об отсутствии гетероскедастичности принимается.

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

3. Автокорреляция. Суть и причины автокорреляции. Последствия и обнаружение автокорреляции. Метод устранения автокорреляции.

Автокорреляция. Суть и причины автокорреляции

Важной предпосылкой построения качественной регрессионной модели по МНК является независимость значений случайных отклонений от значений отклонений во всех других наблюдениях.

Отсутствие зависимости гарантирует отсутствие коррелированности между любыми отклонениями и, в частности, между соседними отклонениями.

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные данные). Автокорреляция остатков (отклонений) обычно встречается в регрессионном анализе при использовании данных временных рядов. При использовании перекрестных данных наличие автокорреляции (пространственной корреляции) крайне редко. В эконометрических задачах значительно чаще встречается так называемая *положительная автокорреляция,* нежели *отрицательная автокорреляция*.

В большинстве случаев положительная автокорреляция вызывается направленным постоянным воздействием некоторых не учтенных в модели факторов.

Суть автокорреляции можно пояснить следующим примером. Пусть исследуется спрос Y на прохладительные напитки в зависимости от дохода X по ежемесячным данным. Трендовая зависимость, отражающая увеличение спроса с ростом дохода, может быть представлена линейной функцией $\hat{y} = a + bx + \varepsilon$, изображенной на рис. 4.3.

Однако фактические точки наблюдений обычно будут превышать трендовую линию в летние периоды и будут ниже нее в зимние.

Аналогичная картина может иметь место в макроэкономическом анализе с учетом циклов деловой активности.

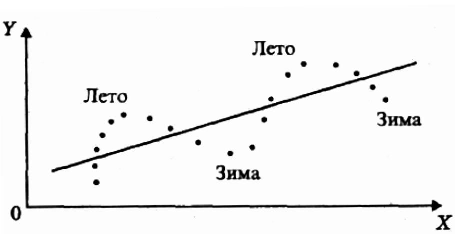


Рис. 4.3

Отрицательная автокорреляция фактически означает, что за положительным отклонением следует отрицательное, и наоборот. Возможная схема рассеивания точек в этом случае представлена на рис. 4.4.

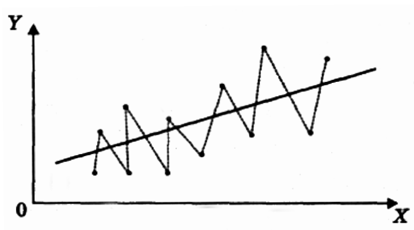


Рис. 4.4

Такая ситуация может иметь место, например, если ту же зависимость между спросом на прохладительные напитки и доходами рассматривать по сезонным данным (зима – лето).

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

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 обусловить автокорреляцию. Проиллюстрируем это следующим примером. Анализируется зависимость предельных издержек *MC* от объема выпуска *Q*. Если для ее описания вместо реальной квадратичной модели выбрать линейную модель, 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.

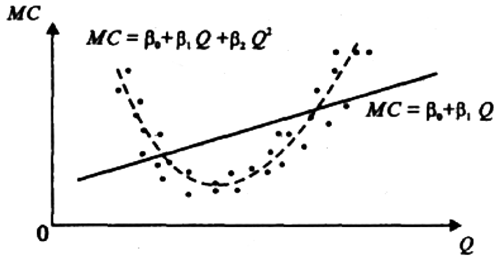


Рис. 4.5

Инерция. Многие экономические показатели (например, инфляция, безработица, ВНП и т. п.) обладают определенной цикличностью, связанной с волнообразностью деловой активности. Действительно, экономический подъем приводит к росту занятости, сокращению инфляции, увеличению ВНП 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 вспять рассматриваемых показателей. В любом случае эта трансформация происходит не мгновенно, а обладает определенной инертностью.

Эффект паутины. Во многих производственных и других сферах 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 прошедшем году вызовет (скорее всего) ее перепроизводство в текущем году, а следовательно, цена на нее снизится и т. д.

Сглаживание данных. Зачастую данные по некоторому продолжительному временному периоду получают усреднением данных по составляющим его по интервалам. 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. Часто дисперсии, вычисляемые по стандартным формулам, являются заниженными, что влечет за собой увеличение t-статистик. Это может привести к признанию статистически значимыми объясняющие переменные, которые в действительности таковыми могут и не являться.

3. Оценка дисперсии регрессии является смещенной оценкой истинного значения, 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

В силу неизвестности значений параметров уравнения регрессии неизвестными будут также и истинные значения отклонений. Поэтому выводы об их независимости осуществляются на основе оценок, полученных из эмпирического уравнения регрессии. Рассмотрим возможные методы определения автокорреляции.

Графический метод

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

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.

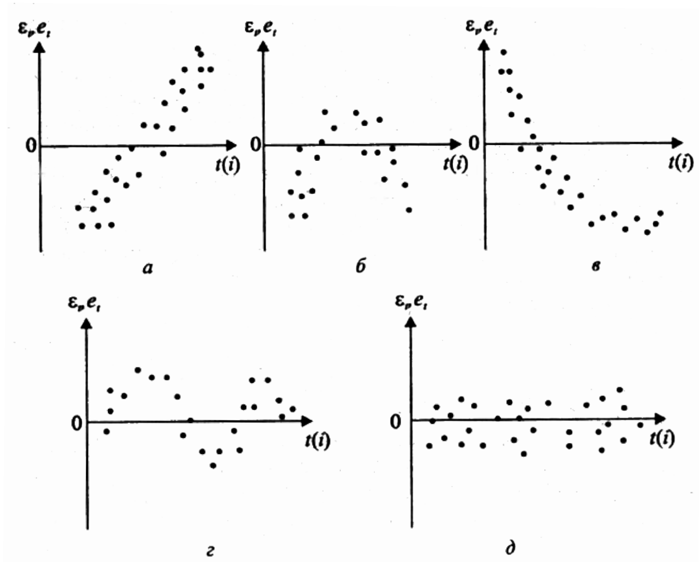


Рис. 4.6

Например, на рис. 4.6б отклонения вначале в основном отрицательные, затем положительные, потом снова отрицательные. Это свидетельствует о наличии между отклонениями определенной зависимости. Более того, можно утверждать, что в этом случае имеет место положительная автокорреляция остатков. Она становится весьма наглядной, если график 4.6б дополнить графиком зависимости e_i от e_{i-1} (рис. 4.7).

Подавляющее большинство точек на этом графике расположено в I и II четвертях декартовой системы координат, подтверждая положительную зависимость между соседними отклонениями.

Следует заметить, что в современных компьютерных прикладных программах для решения задач по эконометрике аналитическое выражение регрессии дополняется графическим представлением результатов. На график реальных колебаний зависимой переменной накладывается график колебаний переменной по уравнению регрессии. Сопоставив эти два графика, можно выдвинуть гипотезу о наличии автокорреляции остатков. Если эти графики пересекаются редко, то можно предположить наличие положительной автокорреляции остатков.

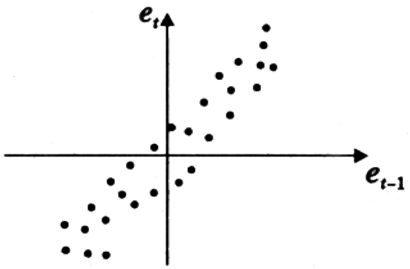


Рис. 4.7

Аналитические методы обнаружения автокорреляции

Наиболее известным аналитическим методом обнаружения автокорреляции первого порядка является критерий Дарбина–Уотсона.

Подробно этот критерий рассмотрен в пункте 4 лекции 6.

4. Обобщенная линейная модель множественной регрессии и обобщенный метод наименьших квадратов

При нарушении гомоскедастичности и наличии автокорреляции ошибок рекомендуется заменять традиционный метод наименьших квадратов обобщенным методом.

Обобщенный метод наименьших квадратов применяется к преобразованным данным и позволяет получать оценки, которые обладают не только свойством несмещенности, но и имеют меньшие выборочные дисперсии. Специфика обобщенного МНК применительно к корректировке данных при автокорреляции остатков будет рассмотрена далее. В этом разделе остановимся на использовании обобщенного МНК для корректировки гетероскедастичности.

Как и раньше, предположим, что среднее значение остатков равно нулю, а дисперсия их пропорциональна величине K, т. е.

$$\sigma^2_{\varepsilon_i} = \sigma^2 K_i,$$

где $\sigma^2_{\varepsilon_i}$ — дисперсия ошибки при конкретном i-м значении фактора;

σ^2 – постоянная дисперсия ошибки при соблюдении предпосылки о гомоскедастичности остатков; K_i – коэффициент пропорциональности, меняющийся с изменением величины фактора, что и обусловливает неоднородность дисперсии.

При этом предполагается, что σ^2 неизвестна, а в отношении величины K выдвигаются гипотезы, характеризующие структуру гетероскедастичности.

В общем виде для уравнения

$$y_i = a + bx_i + \varepsilon_i \text{ при } \sigma^2_{\varepsilon_i} = \sigma^2 K_i$$

модель примет вид:

$$y_i = \alpha + \beta_i x_i + \sqrt{K_i \cdot \varepsilon_i}.$$

В данной модели остаточные величины гетероскедастичны. Предположив в них отсутствие автокорреляции, перейдем к уравнению с гомоскедастичными остатками, поделив все переменные, зафиксированные в ходе i-го наблюдения, на $K^{\frac{1}{2}}$. Тогда дисперсия остатков будет величиной постоянной. Иными словами, от регрессии y по x мы перейдем к регрессии на новых переменных: $\frac{y}{K^{\frac{1}{2}}}$ и $\frac{x}{K^{\frac{1}{2}}}$.

Уравнение регрессии примет вид:

$$\frac{y_i}{\sqrt{K_i}} = \frac{\alpha}{\sqrt{K_i}} + \beta \cdot \frac{x_i}{\sqrt{K_i}} + \varepsilon_i.$$

Исходные данные для этого уравнения будут иметь вид:

$$y = \begin{vmatrix} \frac{y_1}{\sqrt{K_1}} \\ \frac{y_2}{\sqrt{K_2}} \\ \vdots \\ \frac{y_n}{\sqrt{K_n}} \end{vmatrix}, \quad x = \begin{vmatrix} \frac{y_1}{\sqrt{K_1}} \\ \frac{y_2}{\sqrt{K_2}} \\ \vdots \\ \frac{y_n}{\sqrt{K_n}} \end{vmatrix}.$$

По отношению к обычной регрессии уравнение с новыми, преобразованными переменными представляет собой взвешенную регрессию, в которой переменные y и x взяты с весами $1/K^{1/2}$.

Оценка параметров уравнения с преобразованными переменными дается с помощью обобщенного метода наименьших квадратов, для которого необходимо минимизировать сумму квадратов отклонений вида

$$S = \sum \frac{1}{K_i} (y_i - a - bx_i)^2.$$

Соответственно получим следующую систему нормальных уравнений:

$$\begin{cases} \sum \dfrac{y_i}{K_i} = a\sum \dfrac{1}{K_i} + b\sum \dfrac{1}{K_i}, \\ \sum \dfrac{y_i \cdot x_i}{K_i} = a\sum \dfrac{x_i}{K_i} + \sum \dfrac{x_i^2}{K_i}. \end{cases}$$

Если преобразованные переменные x и y взять в отклонениях от средних уровней, то коэффициент регрессии можно определить как

$$b = \frac{\sum \frac{1}{K} \cdot x \cdot y}{\sum \frac{1}{K} \cdot x^2}.$$

При обычном применении метода наименьших квадратов к уравнению линейной регрессии для переменных в отклонениях от средних уровней коэффициент регрессии b определяется по формуле

$$b = \frac{\sum (xy)}{\sum x^2}.$$

Как видим, при использовании обобщенного МНК с целью корректировки гетероскедастичности коэффициент регрессии b представляет собой взвешенную величину по отношению к обычному методу наименьших квадратов с весами 1/K.

Аналогичный подход возможен не только для уравнения парной, но и для уравнения множественной регрессии. Предположим, что рассматривается модель вида

$$y = a + b_1x_1 + b_2x_2 + \varepsilon,$$

для которой дисперсия остатков оказалась пропорциональна K^2, K^2, представляет собой коэффициент пропорциональности, принимающий различные значения для соответствующих значений факторов x_1 и x_2. Ввиду того, что

$$\sigma_{\varepsilon_i}^2 = \sigma^2 K_i^2,$$

рассматриваемая модель примет вид:

$$y_i = a + b_1x_1 + b_2x_2 + K_i\varepsilon,$$

где остатки гетероскедастичны.

Для того чтобы получить уравнение, где остатки ε_i – гомоскедастичны, перейдем к новым, преобразованным переменным, разделив все члены исходного уравнения на коэффициент пропорциональности K_j. Уравнение с преобразованными переменными составит:

$$\frac{y_i}{K_i} = \frac{\alpha}{K_i} + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i.$$

Это уравнение не содержит свободного члена. Вместе с тем, найдя переменные в новом, преобразованном виде и применив к ним обычный МНК, получим иную спецификацию модели

$$\frac{y_i}{K_i} = A + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i.$$

Параметры такой модели зависят от гипотезы, принятой для коэффициентов пропорциональности K. В эконометрических исследованиях довольно часто выдвигается гипотеза, что остатки пропорциональны значению фактора:

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + E,$$

или

$$\frac{y}{x_1} = b_1 + b_2 \frac{x_2}{x_1} + \ldots + b_p \frac{x_p}{x_1} + \varepsilon.$$

Если предположить, что ошибки пропорциональны, то модель примет вид:

$$\frac{y}{x_p} = b_p + b_1 \frac{x_1}{x_p} + \ldots + b_{p-1} \frac{x_{p-1}}{x_p} + \varepsilon.$$

Применение в этом случае обобщенного МНК приводит к тому, что наблюдения с меньшими значениями преобразованных переменных x/K имеют относительно больший вес при определении параметров регрессии, чем с первоначальными переменными. Вместе с тем следует иметь в виду, что новые, преобразованные переменные получают новое экономическое содержание и их регрессия имеет иной смысл, чем регрессия по исходным данным.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов /под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

НЕЛИНЕЙНЫЕ МОДЕЛИ РЕГРЕССИИ И ЛИНЕАРИЗАЦИЯ

План:
1. Некоторые виды нелинейных моделей и их линеаризация.
2. Коэффициент эластичности.
3. Индекс корреляции, индекс детерминации.
4. *F*-критерий Фишера.

1. Некоторые виды нелинейных моделей и их линеаризация

Различают два класса нелинейных регрессий:

1. Регрессии нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, например:
- полиномы различных степеней – $\hat{y}_x = a + b \cdot x + c \cdot x^2$, $\hat{y}_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$;
- равносторонняя гипербола – $\hat{y}_x = a + b/x$;
- полулогарифмическая функция – $\hat{y}_x = a + b \cdot \ln x$.

2. Регрессии нелинейные по оцениваемым параметрам, например:
- степенная – $\hat{y}_x = a \cdot x^b$;
- показательная – $\hat{y}_x = a \cdot b^x$;
- экспоненциальная – $\hat{y}_x = e^{a+b \cdot x}$.

Регрессии нелинейные по включенным переменным приводятся к линейному виду простой заменой переменных (*линеаризация*), а дальнейшая оценка параметров производится с помощью метода наименьших квадратов линейного уравнения.

Рассмотрим некоторые функции.

Парабола второй степени $\hat{y}_x = a + b \cdot x + c \cdot x^2$ приводится к линейному виду с помощью замены: $x = x_1$, $x^2 = x_2$. В результате приходим к двухфакторному уравнению $\hat{y}_x = a + b \cdot x_1 + c \cdot x_2$, оценка параметров которого при помощи МНК приводит к системе следующих нормальных уравнений:

$$\begin{cases} a \cdot n + b \cdot \sum x_1 + c \cdot \sum x_2 = \sum y, \\ a \cdot \sum x_1 + b \cdot \sum x_1^2 + c \cdot \sum x_1 \cdot x_2 = \sum x_1 \cdot y, \\ a \cdot \sum x_2 + b \cdot \sum x_1 \cdot x_2 + c \cdot \sum x_2^2 = \sum x_2 \cdot y. \end{cases}$$

а после обратной замены переменных получим:

$$\begin{cases} a \cdot n + b \cdot \sum x + c \cdot \sum x^2 = \sum y, \\ a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 = \sum x \cdot y, \\ a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 = \sum x^2 \cdot y. \end{cases} \quad (5.1)$$

Парабола второй степени обычно применяется в случаях, когда в определенном интервале значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная на прямую.

Равносторонняя гипербола $\hat{y}_x = a + b/x$ приводится к линейному уравнению простой заменой: $z = 1/x$. Система линейных уравнений при применении МНК будет выглядеть следующим образом:

$$\begin{cases} a \cdot n + b \cdot \sum \frac{1}{x} = \sum y, \\ a \cdot \sum \frac{1}{x} + b \cdot \sum \frac{1}{x^2} = \sum \frac{1}{x} \cdot y. \end{cases} \quad (5.2)$$

Аналогичным образом приводятся к линейному виду зависимости $\hat{y}_x = a + b \cdot \ln x$, $\hat{y}_x = a + b \cdot \sqrt{x}$ и другие.

Несколько иначе обстоит дело с регрессиями нелинейными по оцениваемым параметрам, которые делятся на два типа: *нелинейные модели внутренне линейные* (приводятся к линейному виду с помощью соответствующих преобразований, например логарифмированием) и *нелинейные модели внутренне нелинейные* (к линейному виду не приводятся).

К внутренне линейным моделям относятся, например, степенная функция – $\hat{y}_x = a \cdot x^b$, показательная – $\hat{y}_x = a \cdot b^x$, экспоненциальная – $\hat{y}_x = e^{a + b \cdot x}$, логистическая – $\hat{y}_x = \dfrac{a}{1 + b \cdot e^{-c \cdot x}}$, обратная – $\hat{y}_x = \dfrac{1}{a + b \cdot x}$.

К внутренне нелинейным моделям можно, например, отнести следующие модели: $\hat{y}_x = a + b \cdot x^c$, $\hat{y}_x = a \cdot \left(1 - \dfrac{1}{1 - x^b}\right)$.

Среди нелинейных моделей наиболее часто используется степенная функция $y = a \cdot x^b \cdot \varepsilon$, которая приводится к линейному виду логарифмированием:

$Y = A + b \cdot X + \mathrm{E}$,

где $Y = \ln y$, $X = \ln x$, $A = \ln a$, $\mathrm{E} = \ln \varepsilon$. Применив МНК к преобразованным данным получим:

$$\begin{cases} A \cdot n + b \cdot \sum X = \sum Y, \\ A \cdot \sum X + b \cdot \sum X^2 = \sum X \cdot Y. \end{cases}$$

Из этой системы нормальных уравнений находится промежуточное линейное уравнение, а затем потенцированием находим искомое уравнение.

Широкое использование степенной функции связано с тем, что параметр *b* в ней имеет четкое экономическое истолкование – он является коэффициентом эластичности.

2. Коэффициент эластичности

Коэффициент эластичности показывает, на сколько процентов изменится в среднем результат, если фактор изменится на 1%. Формула для расчета коэффициента эластичности имеет вид:

$$\acute{Y} = f'(x) \cdot \frac{x}{y}. \qquad (5.3)$$

Так как для остальных функций кроме степенной коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора *x*, то обычно рассчитывается средний коэффициент эластичности:

$$\bar{Y} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}} \cdot \qquad (5.4)$$

Формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии представлены в таблице 5.1.

3. Индекс корреляции, индекс детерминации

Уравнение нелинейной регрессии, так же как и в случае линейной зависимости, дополняется показателем тесноты связи. В данном случае это индекс корреляции:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2_y}}. \quad (5.5)$$

Величина данного показателя находится в пределах: $0 \le \rho_{xy} \le 1$. Чем ближе значение индекса корреляции к единице, тем теснее связь рассматриваемых признаков, тем более надежно уравнение регрессии.

Квадрат индекса корреляции называется индексом детерминации, он характеризует долю дисперсии результативного признака *y*, объясняемую регрессией, в общей дисперсии результативного признака:

$$\rho^2_{xy} = 1 - \frac{\sigma^2_{ост}}{\sigma^2_y} = \frac{\sigma^2_{объясн}}{\sigma^2_y}, \quad (5.6)$$

т. е. имеет тот же смысл, что и в линейной регрессии.

Индекс детерминации ρ^2_{xy} можно сравнивать с коэффициентом детерминации r^2_{xy} для обоснования возможности применения линейной функции. Чем больше кривизна линии регрессии, тем величина r^2_{xy} меньше ρ^2_{xy}, а близость этих показателей указывает на то, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию.

Таблица 5.1

Вид функции, *y*	Первая производная, y'	Средний коэффициент эластичности, $\bar{Э}$	Линеаризация
1	2	3	4
	b	$\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$	–
		$\frac{(b + 2c \cdot \bar{x}) \cdot \bar{x}}{a + b \cdot \bar{x} + c \cdot \bar{x}^2}$	$X_1=x$, $X_2=x^2$
		$-\frac{b}{a \cdot \bar{x} + b}$	$X=1/x, Y=y$
		b	$X=\ln x, Y=\ln y$
		$\bar{x} \cdot \ln b$	$X=x, Y=\ln y$
		$\frac{b}{a + b \cdot \ln \bar{x}}$	$X=\ln x, Y=y$
		$\bar{x} \cdot b$	$X=x, Y=\ln y$
$y = \frac{a}{1 + b \cdot e^{-c \cdot x + \varepsilon}}$	$\frac{a \cdot b \cdot c \cdot e^{-c \cdot x}}{(1 + b \cdot e^{-c \cdot x})^2}$	$\frac{b \cdot c \cdot \bar{x}}{b + e^{c \cdot \bar{x}}}$	
		$-\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$	$X=x, Y=1/y$

4. *F*-критерий Фишера

Индекс детерминации используется для проверки существенности в целом уравнения регрессии по *F*-критерию Фишера, для этого рассчитывается величина

$$F = \frac{\rho_{xy}^2}{1-\rho_{xy}^2} \cdot \frac{n-m-1}{m}, \qquad (5.7)$$

где ρ_{xy}^2 – индекс детерминации, n – число наблюдений, m – число параметров при переменной x. Фактическое значение *F*-критерия (5.7) сравнивается с табличным при уровне значимости α и числе степеней свободы $k_2 = n - m - 1$ (для остаточной суммы квадратов) и $k_1 = m$ (для факторной суммы квадратов).

О качестве нелинейного уравнения регрессии можно также судить и по средней ошибке аппроксимации, которая, так же как и в линейном случае, вычисляется по формуле (2.16).

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

3. Айвазян, С. А. Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001. – 300 с.

4. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва: Дело, 1997. – 442 с.

5. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 324 с.

6. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

Лекция 6

МОДЕЛИ ВРЕМЕННЫХ РЯДОВ

План:

1. Временные ряды. Основные элементы временного ряда.
2. Автокорреляция уровней временного ряда.
3. Модели временных рядов.
4. Автокорреляция в остатках. Критерий Дарбина–Уотсона.

1. Временные ряды. Основные элементы временного ряда

При построении эконометрических моделей используются два типа данных:

1. Данные, характеризующие совокупность различных объектов в определенный момент времени.
2. Данные, характеризующие один объект за ряд последовательных моментов времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные на основе второго типа данных, называются *моделями временных рядов*.

Временной ряд (ряд динамики) – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

1) факторы, формирующие тенденцию ряда;
2) факторы, формирующие циклические колебания ряда;
3) случайные факторы.

Соответственно, временной ряд содержит следующие компоненты:

1) *трендовую (T),* описывающую общее изменение со временем результативного признака;
2) *сезонную (S),* отражающую повторяемость данных через небольшой промежуток времени;

3) *случайную (Е),* отражающую влияние случайных факторов.

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 состоит в выявлении и 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.

2. Автокорреляция уровней временного ряда

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

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^{n}\left(y_t - \bar{y}_1\right)\left(y_{t-1} - \bar{y}_2\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \bar{y}_1\right)^2 \sum_{t=2}^{n}\left(y_{t-1} - \bar{y}_2\right)^2}}, \tag{6.1}$$

где

$$\bar{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t, \qquad \bar{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}.$$

Эту величину называют *коэффициентом автокорреляции* уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда y_t и y_{t-1}.

Коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}} \quad (6.2)$$

где

$$\bar{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t, \qquad \bar{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}.$$

Аналогично можно определить коэффициенты автокорреляции более высоких порядков.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, то ряд содержит циклические колебания с периодичностью в τ моментов времени.

Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Свойства коэффициента автокорреляции:

1. 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.

2. 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.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*.

Автокорреляционная функция. Коррелограмма

Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют *автокорреляционной функцией* временного ряда.

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

3. Модели временных рядов.

Моделирование тенденции временного ряда

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда ряда. Этот способ называют *аналитическим выравниванием временного ряда*.

Для построения трендов чаще всего применяются следующие функции: линейный тренд: $\hat{y}_t = a + b \cdot t$; гипербола: $\hat{y}_t = a + b/t$; экспоненциальный тренд: $\hat{y}_t = e^{a+b \cdot t}$ (или $\hat{y}_t = a \cdot b^t$); степенная функция: $\hat{y}_t = a \cdot t^b$; полиномы различных степеней:

$$\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + \ldots + b_m \cdot t^m \cdot$$

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время $t = 1, 2, \ldots, n$, а в качестве зависимой переменной – фактические уровни временного ряда $\hat{y}_t$. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Наиболее простую экономическую интерпретацию имеет линейная функция $y_t = a + bt$: a – начальный уровень временного ряда в момент времени $t = 0$; b – средний за период абсолютный прирост уровней ряда.

Параметры a и b находятся по формулам:

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} t_i}{n}, \qquad b = \frac{n\sum_{i=1}^{n} t_i y_i - \sum_{i=1}^{n} t_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} t_i^2 - (\sum_{i=1}^{n} t_i)^2}.$$

Существует несколько способов определения типа тенденции. К числу наиболее распространенных способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни $\hat{y}_t$ и $\hat{y}_{t-1}$ тесно коррелируют.

В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

Моделирование сезонных колебаний

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий:

$$Y = T + S + E\,. \tag{6.3}$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (T), сезонной (S) и случайной (E) компонент.

Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E\,. \tag{6.4}$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой (T), сезонной (S) и случайной (E) компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T, S и E для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги:

1. Выравнивание исходного ряда методом скользящей средней.
2. Расчет значений сезонной компоненты S.
3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных ($T + E$) в аддитивной или ($T \cdot E$) в мультипликативной модели.
4. Аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений T с использованием полученного уравнения тренда.
5. Расчет полученных по модели значений ($T + S$) или ($T \cdot S$).
6. Расчет абсолютных и/или относительных ошибок.

Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок E для анализа взаимосвязи исходного ряда и других временных рядов.

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле.

Скорректированные значения сезонной компоненты в аддитивной модели равны $S_i = \overline{S}_i - k$, где $k = \dfrac{\sum_{i=1}^{n} S_i}{n}$, в мультипликативной модели S_i получаются при умножении ее средней оценки $\overline{S}_i$ на корректирующий коэффициент k, где $k = \dfrac{n}{\sum_{i=1}^{n} S_i}$.

Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент, в мультипликативной модели есть произведение трендовой и сезонной компонент.

4. Автокорреляция в остатках. Критерий Дарбина—Уотсона

Рассмотрим уравнение регрессии вида

$$y_t = a + b_1 x_1 + b_2 x_2 + ... + b_m x_m + \mathring{a}_t,$$

где m — число независимых переменных модели.

Для каждого момента (периода) времени $t = 1$ n-значение компоненты ε_t определяется как

$$\varepsilon_t = y_t - \hat{y}_t.$$

Рассматривая последовательность остатков как временной ряд, можно построить график их зависимости от времени. В соответствии с предпосылками МНК остатки ε_t должны быть случайными. Однако при моделировании временных рядов нередко встречается ситуация, когда остатки содержат тенденцию или циклические колебания. Это свидетельствует о том, что каждое следующее значение остатков зависит от предшествующих. В этом случае говорят о наличии автокорреляции остатков.

Автокорреляция в остатках может быть вызвана несколькими причинами, имеющими различную природу:

1. Она может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака.

2. В ряде случаев автокорреляция может быть следствием неправильной спецификации модели, модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t.

Одним из наиболее распространенных методов определения автокорреляции в остатках — это расчет критерия Дарбина–Уотсона:

$$d = \frac{\sum_{t=2}^{n}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{t=1}^{n}\varepsilon_t^2}. \tag{6.5}$$

Величина d есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии.

Можно показать, что при больших значениях n существует следующее соотношение между критерием Дарбина–Уотсона d и коэффициентом автокорреляции остатков первого порядка r_1: $d = 2 \cdot \left(1 - r_1\right)$.

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1 = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1 = 0$ и $d = 2$. Т.е. $0 \le d \le 4$.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина–Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам определяются критические значения критерия Дарбина–Уотсона d_Lи d_U для заданного числа наблюдений n, числа независимых переменных модели m и уровня значимости α. По этим значениям числовой промежуток [0; 4] разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью $1-\alpha$ осуществляется следующим образом:

$0 < d < d_L$ – есть положительная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1-\alpha$ принимается H_1;

$d_L < d < d_U$ – зона неопределенности;

$d_U < d < 4 - d_U$ – нет оснований отклонять H_0, т. е. автокорреляция остатков отсутствует;

$4 - d_U < d < 4 - d_L$ – зона неопределенности;

$4 - d_L < d < 4$ – есть отрицательная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1-\alpha$ принимается H_1^*.

Если фактическое значение критерия Дарбина–Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / под ред. проф. Н.Ш.Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

3. Айвазян, С. А. Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001. – 300 с.

4. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва: Дело, 1997. – 442 с.

5. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. –Москва: Финансы и статистика, 2005. – 324 с.

6. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

НЕСТАЦИОНАРНЫЕ ВРЕМЕННЫЕ РЯДЫ

План:

1. Модели нестационарных временных рядов.

2. Модель авторегрессии – проинтегрированного скользящего среднего.

1. Модели нестационарных временных рядов

В моделях временных рядов, рассмотренных выше, остатки стационарны, а нестационарность самих рядов x_t и y_t обуславливалась наличием неслучайной компоненты (тренда). После выделения тренда все ряды оказывались стационарными, причем стационарность считалась заранее известной, априорной. На практике, однако, такая ситуация редко имеет место. Между тем включение в модель нестационарных рядов может привести к совершенно неверным результатам. В частности, стандартный анализ с помощью метода наименьших квадратов модели

$$y_t = \alpha + \beta x_t + \varepsilon$$

может показать наличие существенной значимости коэффициента β даже в том случае, если величины x_t и y_t, являются независимыми.

Такое явление носит название *ложной регрессии* и имеет место именно в том случае, когда в модели используются нестационарные временные ряды.

Таким образом, при моделировании какой-либо зависимости между величинами x_t и y_t естественно возникает вопрос: можно ли считать соответствующие временные ряды стационарными?

В рамках этой лекции мы рассмотрим некоторые вопросы, связанные с *нестационарными временными рядами*. При этом мы не будем ставить задачу обстоятельного изложения теории нестационарных рядов, так как это потребовало бы использования математического аппарата, существенно выходящего за рамки нашего рассмотрения. Поэтому мы ограничимся лишь тем, что затронем основные проблемы, возникающие при эконометрическом моделировании нестационарных временных рядов.

Итак, пусть имеется временной ряд y_t, при этом мы считаем, что в нем отсутствует неслучайная составляющая. Для простоты также будем считать, что среднее его значение равно нулю (очевидно, ряд остатков регрессионной модели удовлетворяет этим условиям).

Если ряд является стационарным, то в каждый следующий момент времени его значение «стремится вернуться к нулевому среднему». Иными словами, если мы будем объяснять значение y_t предыдущим значением y_{t-1}, то объясненная часть y_t будет находиться ближе к нулю, чем значение y_{t-1}.

Математически строго это условие можно сформулировать следующим образом: рассмотрим регрессионную модель

$$y_t = \rho y_{t-1} + \xi_t.$$

Истинное значение параметра р должно удовлетворять условию $\rho < 1$. В случае, если $\rho = 1$, мы имеем ситуацию, когда последующее значение одинаково легко может как приближаться к нулевому среднему, так и отдаляться от него. Соответствующий случайный процесс называется «случайным блужданием». Очевидно, дисперсия в этом случае растет. В самом деле, имеем:

$$D(y_t) = D(y_{t-1}) + \sigma_t^2,$$

т. е. дисперсия D неограниченно возрастает, а значит, ряд y_t не является стационарным.

Разумеется, в случае $|\rho| > 1$ ряд тем более не будет стационарным, значения его стремительно нарастают. Соответствующий процесс иногда называется *взрывным*. Однако в реальных экономических задачах он никогда не возникает.

Практика показывает, что чаще всего в эконометрических исследованиях нестационарность рассматриваемого временного ряда носит именно характер случайного блуждания. Таким образом, вопрос о нестационарности ряда y_t, как правило, сводится к следующему: верно ли, что в регрессии $y = \rho y_{t-1} + \xi$ истинное значение параметра ρ равно единице? Соответствующая задача называется проблемой *единичного корня*.

2. Модель авторегрессии – проинтегрированного скользящего среднего

Итак, пусть имеется временной ряд y_t. Рассмотрим модель авторегрессии

$$y_t = \rho y_{t-1} + \xi_t.$$

Будем предполагать, что ошибки регрессии ξ_t *независимы и одинаково распределены, т. е. образуют белый шум.* Переходя к разностным величинам, получим:

$$\Delta y_t = \lambda\ y_{t-1} + \xi_t$$

где $\Delta y_t = y_t - y_{t-1}$, $\lambda = \rho - 1.$

Тогда проблема единичного корня сводится к следующей: верно ли, что в модели $\Delta\ y = \lambda\ y_{t-1} + \xi_t$ истинное значение параметра λ равно нулю?

На первый взгляд кажется, что вопрос может быть решен тестированием гипотезы $\lambda = 0$ с помощью статистики Стьюдента. Однако ситуация оказывается сложнее. В том случае, если ряд y_t на самом деле нестационарный, т. е. если на самом деле $\lambda = 0$, стандартная t-статистика не имеет распределения Стьюдента.

Распределение *t*-статистики в этом случае описано Дики и Фуллером. Ими же получены критические значения для отвержения гипотезы о нестационарности ряда. Они существенно отличаются от критических значений распределения Стьюдента. В результате оказывается, что использование обычного *t*-теста приводит к тому, что гипотеза о нестационарности временного ряда отвергается слишком часто, в том числе и тогда, когда ряд действительно является нестационарным.

Таким образом, проблему единичного корня следует решать с помощью теста Дики–Фуллера, который реализован в большинстве современных регрессионных пакетов. Он является обобщением обычного теста Дики–Фуллера, и в правую часть исследуемого выражения добавляются слагаемые вида $\Delta y_{t-1}.......\Delta y_{t-p}$, то есть тестируется гипотеза $\lambda=0$ для модели

$$\Delta y_t = \lambda y_{t-1} + \lambda_1 \Delta y_{t-1} + ... + \lambda_p \Delta_{t-p} + \varepsilon_t.$$

Это соответствует тому, что рассматривается вместо предыдущего уравнения уравнение

$$y_t = \rho\, y_{t-1} + \rho_2 y_{t-2} + ... + \rho_{p+1} y_{t-p-1},$$

т. е. пытаются идентифицировать ряд как авторегрессионный порядка $p+1$.

Добавление приращений Δy_{t-1} производится для того, чтобы с возможно большей достоверностью избавиться от автокорреляции ошибок. При этом считается, что распределение Дики–Фуллера t-статистики имеет место лишь в том случае, если ошибки являются белым шумом. Однако добавление приращений в правую часть снижает мощность теста Дики–Фуллера.

Что делать в том случае, если ряд y_t оказался нестационарным?

Часто при этом оказывается, что стационарным является ряд приращений Δy_t.

Если ряд Δy_t, является стационарным, то исходный нестационарный ряд y_t называется интегрируемым (или однородным). В более общем случае нестационарный ряд y_t называется *интегрируемым (однородным) k-го* порядка, если после k-кратного перехода к приращениям

$$d^k y_t = d^{k-1} y_t - d^{k-1} y_{t-1},$$

где $dy_t = \Delta y_t$.

Соответственно получается стационарный ряд $d^k y_t$.

Если при этом стационарный ряд $d^k y$ корректно идентифицируется как ARIMA (p,q) (авторегрессионная модель скользящей средней), то нестационарный ряд y обозначается ARIMA (p, q, k). Модель ARIMA (p, q, k) означает модель авторегрессии – проинтегрированной скользящей средней – и известна как модель Бокса–Дженкинса. Эта модель может достаточно успешно описывать поведение *нестационарных временных рядов (в том числе содержащих сезонную и/или циклическую компоненты),* что позволяет эффективно использовать ее в задачах кратко- и среднесрочного автопрогноза. Процедура подбора модели реализована во многих эконометрических пакетах.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

3. Айвазян, С. А. Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001. –300 с.

4. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва: Дело, 1997. – 442 с.

5. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 324 с.

6. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

Лекция 8

СИСТЕМА ЛИНЕЙНЫХ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

План:

1. Система линейных одновременных уравнений.
2. Структурная и приведенная формы модели систем одновременных уравнений.
3. Условия идентифицируемости уравнений системы.
4. Индентификация систем уравнений.

1. Система линейных одновременных уравнений

В экономических исследованиях важное место занимает проблема описания структуры связей между переменными системой так называемых одновременных уравнений, которые называются также структурными уравнениями.

В общем случае система уравнений в эконометрических исследованиях может быть построена по-разному. Рассмотрим возможные варианты построения такого рода систем.

Система независимых уравнений

Система независимых уравнений возможна, когда каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \tag{8.1}$$

Для нахождения его параметров используется метод наименьших квадратов. По существу, каждое уравнение этой системы является уравнением регрессии. Так как фактические значения зависимой переменной отличаются от теоретических на величину случайной ошибки, то в каждом уравнении присутствует величина случайной ошибки ε_i.

Рекурсивные системы одновременных уравнений

Если зависимая переменная y одного уравнения выступает в виде фактора x в другом уравнении, то исследователь может строить модель в виде *системы рекурсивных уравнений:*

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + \ldots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \tag{8.2}$$

В данной системе зависимая переменная y включает в каждое последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений наряду с набором собственно факторов x. Каждое уравнение этой системы может рассматриваться самостоятельно, и его параметры определяются методом наименьших квадратов (МНК).

Системы одновременных уравнений. Модель спроса и предложений как пример системы одновременных уравнений

Наибольшее распространение в эконометрических исследованиях получила *система взаимозависимых уравнений*. В ней одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях — в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1m}y_m \qquad + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2m}y_m \qquad + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + \ldots + b_{3m}y_m \qquad + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + b_{m2}y_2 + \ldots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_n. \end{cases} \tag{8.3}$$

Система взаимозависимых уравнений получила название *системы совместных, одновременных уравнений*. В эконометрике эта система уравнений называется также *структурной формой модели*.

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

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

Эндогенные переменные – это зависимые переменные, число которых равно числу уравнений в системе и которые обозначаются через y, взаимозависимые переменные, которые определяются внутри модели (системы).

Экзогенные переменные – это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них, независимые переменные, которые определяются вне системы. Обозначаются через x.

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

Предопределенными переменными называются экзогенные и лаговые эндогенные переменные системы.

Классическим примером системы одновременных уравнений является система уравнений одновременного формирования спроса Q^d и предложения Q^s товара в зависимости от его цены P:

$$Q^d = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_1,$$

$$Q^s = \beta_4 + \beta_5 P + \varepsilon_2,$$

где I – доход.

Если предположить, что рынок находится в состоянии равновесия, то в этих равенствах следует положить $Q^d = Q^s = P$. В этом случае наблюдаемое значение P – это цена равновесия, которая формируется одновременно со спросом и предложением. Таким образом, мы должны считать P и Q объясняемыми переменными, а величину дохода I – объясняющей переменной.

2. Структурная и приведенная формы модели систем одновременных уравнений

Структурная форма модели позволяет увидеть влияние изменений любой экзогенной переменной на значения эндогенной переменной.

Структурная форма модели в правой части содержит при эндогенных переменных коэффициенты b_{ik}, а при экзогенных переменных – коэффициенты a_{ij}, которые называются *структурными коэффициентами* модели. Все переменные в модели выражены в отклонениях от среднего уровня, т. е. под x подразумевается $x - \overline{x}$, а под y – соответственно $y - \overline{y}$. Поэтому свободный член в каждом уравнении системы (3.3) отсутствует.

Использование МНК для оценивания структурных коэффициентов модели дает, как принято считать в теории, смещенные и несостоятельные оценки. Поэтому обычно для определения структурных коэффициентов модели структурная форма модели преобразуется в *приведенную форму* модели.

Приведенная форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1n}x_n + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2n}x_n + u_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = \delta_{m1}x_1 + \delta_{m2}x_2 + \ldots + \delta_{mn}x_n + u_m, \end{cases} \tag{8.4}$$

где δ_{ij} – коэффициенты приведенной формы модели, u_i – остаточная величина для приведенной формы.

По своему виду приведенная форма модели ничем не отличается от системы независимых уравнений, параметры которой оцениваются традиционным МНК. Применяя МНК к этой форме модели, можно оценить δ_{ij}, а затем оценить значения эндогенных переменных через экзогенные.

Необходимо отметить, что коэффициенты приведенной формы модели представляют собой нелинейные функции коэффициентов структурной формы модели.

3. Условия идентифицируемости уравнений системы

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

С позиции идентифицируемости структурные модели можно подразделить на три вида:

1) идентифицируемые;

2) неидентифицируемые;

3) сверхидентифицируемые.

Модель *идентифицируема*, если все структурные ее коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель *неидентифицируема*, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы. Сверхидентифицируемая модель, в отличие от неидентифицируемой модели, практически решаема, но требует для этого специальных методов исчисления параметров.

Структурная форма модели всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

4. Идентификация систем уравнений

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы.

Необходимое условие идентификации

Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в i-м уравнении системы через H, а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, – через D, то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

Таблица 8.1

$D+1=H$	уравнение идентифицируемо
$D+1<H$	уравнение неидентифицируемо
$D+1>H$	уравнение сверхидентифицируемо

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Если необходимое условие выполнено, то далее проверяется достаточное условие идентификации.

Достаточное условие идентификации

Уравнение идентифицируемо, если определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен нулю и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

3. Айвазян, С. А. Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001. – 300 с.

4. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва: Дело, 1997. – 442 с.

5. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 324 с.

6. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

Лекция 9

МЕТОДЫ ОЦЕНКИ ПАРАМЕТРОВ СТРУКТУРНОЙ ФОРМЫ МОДЕЛИ

План:
1. Методы оценки параметров структурной формы модели.
2. Косвенный метод наименьших квадратов.
3. Двухшаговый метод наименьших квадратов.
4. Трехшаговый метод наименьших квадратов.

1. Методы оценки параметров структурной формы модели

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Наибольшее распространение в литературе получили следующие методы оценивания коэффициентов структурной модели:

1) косвенный метод наименьших квадратов (КМНК);
2) двухшаговый метод наименьших квадратов (ДМНК);
3) трехшаговый метод наименьших квадратов (ТМНК);
4) метод максимального правдоподобия с полной информацией;
5) метод максимального правдоподобия при ограниченной информации.

Рассмотрим вкратце сущность каждого из этих методов.

Косвенный и двухшаговый методы наименьших квадратов подробно рассматриваются как традиционные методы оценки коэффициентов структурной модели. Эти методы достаточно легко реализовать.

Косвенный метод наименьших квадратов применяется для идентифицируемой системы одновременных уравнений.

Двухшаговый метод наименьших квадратов используется для оценки коэффициентов сверхидентифицируемой модели.

Метод максимального правдоподобия рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК. Однако при большом числе уравнений системы этот метод приводит к достаточно сложным вычислительным процедурам. Поэтому в качестве модификации используется метод максимального правдоподобия при ограниченной информации (метод наименьшего дисперсионного отношения). В отличие от метода максимального правдоподобия, в данном методе сняты ограничения на параметры, связанные с функционированием системы в целом. Это делает решение более простым, но трудоемкость вычислений остается достаточно высокой. Несмотря на его популярность он был практически вытеснен двухшаговым методом наименьших квадратов в связи с гораздо большей простотой последнего. Этому способствовала также разработка семейства оценок коэффициентов структурной модели.

Дальнейшим развитием двухшагового метода наименьших квадратов является трехшаговый МНК. Этот метод оценивания пригоден для всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным оказывается двухшаговый метод наименьших квадратов.

2. Косвенный метод наименьших квадратов

Косвенный метод наименьших квадратов (КМНК) применяется в случае точно идентифицируемой структурной модели.

Процедура применения КМНК предполагает выполнение следующих этапов работы:

1. Структурная модель преобразовывается в приведенную форму модели.
2. Для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты δ_{ij} .
3. Коэффициенты приведенной формы модели трансформируются в параметры структурной модели.

Оценка значимости модели дается через F-критерий и R^2 для каждого уравнения в отдельности.

При сравнении результатов, полученных традиционным методом наименьших квадратов и с помощью косвенного метода наименьших квадратов, следует иметь в виду, что традиционный МНК, применяемый к каждому уравнению структурной формы модели, взятому в отдельности, дает смещенные оценки структурных коэффициентов.

Так, рассмотрим две взаимосвязанные регрессии:

$$\begin{cases} y = ax + \varepsilon_1, \\ x = by + \varepsilon_2. \end{cases}$$

В этом случае коэффициент регрессии отличается от структурного коэффициента и совпадает с ним только в одном частном случае, когда переменная y не содержит ошибок (т. е. $\varepsilon_1 = 0$), а ошибки переменной x имеют дисперсию, равную единице.

Кроме того, при интерпретации коэффициентов множественной регрессии предполагается независимость факторов друг от друга, что становится невозможным при рассмотрении системы совместных уравнений.

Нарушение предпосылки независимости факторов друг от друга при использовании традиционного МНК в системе одновременных уравнений приводит к несостоятельности оценок структурных коэффициентов, в ряде случаев они оказываются экономически бессмысленными. Опасность таких результатов возрастает при увеличении числа эндогенных переменных в правой части системы, потому что становится невозможным расщепить совместное влияние эндогенных переменных и видеть изолированные меры их воздействия в соответствии с предпосылками традиционного метода наименьших квадратов.

Компьютерная программа применения КМНК предполагает, что система уравнений содержит в правой части в каждом уравнении как эндогенные, так и экзогенные переменные. Между тем могут быть системы, в которых в одном из уравнений, например, отсутствуют экзогенные переменные. Например, возможна модель экономики страны с четырьмя эндогенными и двумя экзогенными переменными, в которой в первом уравнении системы не содержалось ни одной экзогенной переменной. Для такой модели непосредственное получение структурных коэффициентов невозможно. В этом случае сначала определяют систему приведенной формы модели, решаемой обычным МНК, а затем переходят путем алгебраических преобразований к коэффициентам структурной модели.

3. Двухшаговый метод наименьших квадратов

Если система сверхидентифицируема, то КМНК не используется. Это связано с тем, что он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является двухшаговый метод наименьших квадратов.

Основная идея ДМНК состоит в следующем: на основе приведенной формы модели получить для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения. Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения.

Метод получил название ДМНК в связи с тем, что дважды используется МНК: на первом шаге при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной $\tilde{y}_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \ldots + \delta_{in}x_n$ и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

1) все уравнения системы сверхидентифицируемы;

2) система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется ДМНК.

Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

4. Трехшаговый метод наименьших квадратов

Наиболее эффективная процедура оценивания систем регрессионных уравнений сочетает метод одновременного оценивания и метод инструментальных переменных. Соответствующий метод называется трехшаговым методом наименьших квадратов. Он заключается в том, что на первом шаге к исходной модели применяется обобщенный метод наименьших квадратов с целью устранения корреляции случайных членов. Затем к полученным уравнениям применяется двухшаговый метод наименьших квадратов.

Очевидно, что если случайные члены исходного уравнения не коррелируют, трехшаговый метод сводится к двухшгаговому, в то же время если матрица при эндогенных переменных единичная, трехшаговый метод представляет собой процедуру одновременного оценивания уравнений как внешне не связанных.

Литература:

1. Кремер, Н. Ш., Путко, Б. А. Эконометрика: Учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002. – 311с.

2. Красе, М. С., Чупрынов, Б. П. Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004. – 464 с.

3. Айвазян, С. А. Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001. –300 с.

4. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – Москва: Дело, 1997. – 442 с.

5. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 324 с.

6. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005. – 225 с.

СЛОВАРЬ ТЕРМИНОВ

А

Автокорреляция – явление взаимосвязи между рядами первоначальным и этим же рядом, сдвинутым относительно первоначального положения на h моментов времени.

Автокорреляция уровней временного ряда – корреляционная зависимость межу последовательными уровнями ряда.

Авторегрессионная модель – разновидность динамической эконометрической модели, которая содержит в качестве факторных переменных лаговые значения эндогенных переменных.

Авторегрессия – регрессия, учитывающая влияние предыдущих уровней ряда на последующие.

Адаптивных ожиданий модель – разновидность динамической эконометрической модели, в которой учитывается ожидаемое значение факторного признака x_t+1.

Аддитивная модель временного ряда – модель, в которой все компоненты ряда динамики представлены как сумма этих составляющих $Y = T + S + E$. Ее применяют в случае, когда амплитуда сезонных колебаний со временем не меняется.

Аналитическое выравнивание – динамического ряда проводится при помощи математической формулы, отражающей общую тенденцию ряда.

Б

«Белый шум» – чисто случайный процесс, то есть ряд независимых, одинаково распределенных случайных величин.

Бета-коэффициент – показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение результативного признака при изменении факторного признака на величину своего среднеквадратического отклонения; β-коэффициент линий регрессии характеризует наиболее крупные резервы улучшения изучаемого признака.

Биссериальный коэффициент корреляции – оценивание связи между качественным альтернативным и количественным варьирующим признаками.

В

Верификация модели – проверка истинности модели, определение соответствия построенной модели реальному экономическому явлению.

Временной лаг – сдвиг, временное смещение уровней временного ряда относительно первоначального положения на h моментов времени.

Временной ряд – ряд последовательно расположенных во времени числовых показателей, которые характеризуют уровень состояния и изменения явления или процесса.

Временные данные – набор сведений, характеризующих один и тот же объект за разные периоды времени.

Г

Гармонический анализ – нахождение конечной суммы уровней с использованием функций косинусов и синусов времени.

Графический метод – способ распознавания типа тренда, при котором временные интервалы откладывают на оси абсцисс, величины уровней – по оси ординат. При этом по каждой оси следует установить такой масштаб, чтобы ширина графика была примерно в 1,5 раза больше его высоты.

Гомоскедастичностъ – дисперсия каждого отклонения одинакова для всех значений x.

Гетероскедастичность – непостоянство дисперсий отклонений каждого отклонения для всех значений x.

Гиперболическая модель – модель вида $y = \mathrm{a} + \frac{\mathrm{b}}{\mathrm{x}} + \varepsilon$.

Д

Двухшаговый метод наименьших квадратов – один из способов решения систем одновременных уравнений, который применяется как для идентифицируемых, так и для сверхидентифицируемых моделей.

Динамическая эконометрическая модель – учитывает в данный момент времени значения входящих в нее переменных, относящихся к текущему и к предыдущему моменту времени.

Диаграмма – графическое изображение статистических данных, иллюстрирующее соотношение между сравниваемыми величинами.

Диаграмма рассеяния – графическое изображение корреляционного поля, наглядно иллюстрирующее соотношение между двумя переменными величинами.

Дискретный вариационный ряд – ряд, в котором величина статистического признака может быть представлена только целочисленными, дискретными значениями.

Дисперсия – характеристика вариации признака, выражаемая средним квадратом отклонений индивидуальных значений вариант от средней арифметической по данной совокупности.

Доверительный интервал – границы, в которых с той или иной вероятностью заключена искомая величина по генеральной совокупности.

Долгосрочный мультипликатор – показатель модели авторегрессии, который определяет общее абсолютное изменение результата в долгосрочном периоде.

З

Задача эконометрики – оценка направленных действий на достижение и повышение экономической эффективности, кроме того, задачей эконометрики является прогнозирование путей развития макро- и микроэкономических факторов.

Закон больших чисел – количественная закономерность, подтверждающая, что практически маловероятно значительное отклонение средней арифметической выборочной совокупности от средней арифметической генеральной совокупности, если число наблюдений достаточно велико.

Закрытый интервал – интервал, имеющий верхнюю и нижнюю границы значений группировочного признака.

И

Идентификация модели – проведение статистического анализа модели и оценивания качества ее параметров; установление соответствия между приведенной и структурной формами модели.

Идентифицируемая модель – разновидность структурной модели системы одновременных уравнений, в которой все структурные коэффициенты однозначно определяются через приведенные коэффициенты.

Индекс множественной корреляции – характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, оценивает тесноту совместного влияния факторов на результат.

Индекс сезонности – процентное соотношение фактических внутригодовых уровней и постоянной или переменной средней.

Индекс частной корреляции – характеризует тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Интерполяция – применяется на этапе предварительной обработки данных и предполагает определение значений уровней ряда внутри заданного интервала.

Интервальный ряд динамики – ряд последовательно расположенных показателей за определенный период.

Интерполяция – приближенный расчет уровней, лежащих внутри ряда динамики, но почему-либо неизвестных.

К

Ковариация – характеризует сопряженность вариации двух признаков и представляет собой статистическую меру взаимодействия двух случайных переменных.

Коинтеграция – причинно-следственная связь в уровнях двух или более временных рядов, которая выражается в совпадении или противоположной направленности их тенденций и случайной колеблемости.

Коллинеарные переменные – переменные, находящиеся между собой в линейной зависимости.

Корреляция – это статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Корреляционная зависимость – связь, при которой каждому значению независимой переменной x соответствует определенное математическое ожидание (среднее значение) зависимой переменной *y*.

Корреляционный анализ – заключается в количественном определении тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

Корреляционная связь – изменение среднего значения результативного признака, которое обусловливается изменением факторных признаков.

Корреляционное отношение – показывает связь между двумя признаками.

Косвенный метод наименьших квадратов – один из способов решения систем одновременных уравнений, основанный на получении состоятельных и несмещенных оценок параметров структурной формы модели по оценкам параметров приведенной формы.

Койка метод – оценивание эконометрических моделей с бесконечным числом лагов.

Криволинейная зависимость – это связь, при которой с возрастанием величины факторного признака возрастание (или убывание) результативного признака происходит неравномерно (выражается уравнениями кривых линий).

Коэффициенты ассоциации и контингенции – определяют тесноту связи двух качественных признаков, каждый из которых состоит только из двух групп.

Коэффициент конкордации – множественный коэффициент ранговой корреляции для определения меры тесноты связи между произвольным числом ранжированных значений признака, измеряемых в порядковой шкале.

Коэффициент корреляции – показатель меры тесноты взаимосвязи двух случайных величин.

Коэффициенты корреляции Спирмена и Кендалла – определяют тесноту связи между двумя количественными или качественными признаками после предварительного ранжирования их по возрастанию или убыванию.

Коэффициент опережения (замедления) – относительный показатель, характеризующий сравнение динамических рядов, относящихся к двум пространственным объектам (странам, республикам и т. д.).

Коэффициент ранговой корреляции – показатель меры тесноты связи между предварительно ранжированными значениями качественных признаков.

Коэффициент регрессии – показатель, характеризующий в среднем изменение значения результативного признака при изменении факторного признака на единицу собственного измерения.

Коэффициент детерминации – характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака.

Коэффициент взаимной сопряженности Пирсона–Чуврова – определение тесноты связи двух качественных признаков, каждый из которых состоит более чем из двух групп.

Коэффициент эластичности – показывает, на сколько процентов изменяется результативный признак y при изменении факторного признака x на один процент.

Л

Лаг – промежуток времени отставания одного явления от другого, связанного с ним.

Лаги Алмон – один из видов модели с распределенным лагом, который характеризуется полиномиальной структурой и конечной величиной лага.

Лаговые (экзогенные или эндогенные) – это такие переменные модели, которые датируются предыдущими моментами времени и находятся в уравнении с текущими переменными.

Линейный коэффициент корреляция – показатель меры тесноты и направления связи между двумя коррелируемыми величинами.

Линейная модель – модель вида $y = a + bx + \varepsilon$.

Линейная связь – статистическая связь между явлениями, выраженная уравнением прямой линии.

М

Математическое ожидание – сумма произведений значений случайной величины на соответствующие вероятности.

Метод разности средних двух частей одного и того же ряда – один из критериев проверки на наличие тренда, где проверяется гипотеза о существовании разности средних H_0: $\overline{Y}_1 = \overline{Y}_2$. Для этого временной ряд разбивают на две равные или почти равные части. В качестве критерия проверки гипотезы принимают критерий Стьюдента. Если $t_{фак} \geqslant t_{таб}$, то гипотеза об отсутствии тренда отвергается; если $t_{фак} \leqslant t_{таб}$, то гипотеза H_0 принимается.

Метод Фостера–Стюарта – критерий проверки на наличие тренда, где определяется наличие тенденции явления и тренд дисперсии уровней временного ряда. Часто этот метод используют в случае детального анализа временного ряда и построения по нему прогнозов. Вычисление критерия проводится поэтапно: проводится сравнительная оценка каждого уровня временного ряда со всеми предыдущими уровнями; вычисляют значения величин q и d; определяют критерий Стьюдента и сравнивают его с табличным значением. Величина d характеризует тенденцию изменения средней и имеет два предела: нижний и верхний. Величина q характеризует тенденцию изменения дисперсии временного ряда и принимает значения в пределах: $0 \leqslant q \leqslant n-1$.

Метод наименьших квадратов – метод, согласно которому сумма отклонений фактических значений результативных показателей от теоретических, найденных по уравнению связи, должна быть минимальной.

Механическое сглаживание – метод нахождения плавных уровней ряда динамики путем использования скользящих средних. Различают метод невзвешенных и взвешенных скользящих средних.

Множественная корреляция – это зависимость между результативным признаком и двумя и более факторными признаками, включенными в исследование.

Многофакторная (множественная) зависимость – это связь между несколькими факторными признаками и результативным признаком (факторы действуют комплексно, т. е. одновременно и во взаимосвязи).

Множественная регрессия – характеризует связь между результативным признаком и двумя и более факторными признаками.

Множественный коэффициент корреляции – отражает связь между результативным и несколькими факторными признаками.

Моментный ряд динамики – это ряд последовательно расположенных показателей на определенную дату.

Модель временного ряда – разновидность эконометрической модели, в которой результативный признак является функцией переменной времени или переменных, относящихся к другим моментам времени.

Модель регрессионная с одним равнением – имеет вид $Y \rightarrow Mx(Y) + \acute{\varepsilon}$, где результативный признак является функцией от факторных признаков $Y = f(x_1, x_2, \ldots x_n) + \acute{\varepsilon}$, а объясненная составляющая $f(x_1, x_2, \ldots x_n)$ представляет собой ожидаемое значение результата Y при заданных значениях факторов $x_1, x_2, \ldots x_n$.

Мультиколлинеарность – тесная корреляционная взаимосвязь между отбираемыми для анализа факторами, совместно воздействующими на общий результат.

Мультипликативная модель временного ряда – модель, в которой факторы влияния представлены в виде произведения составляющих $y_t = u_t \cdot v_t \cdot \acute{\varepsilon}_t$.Такую модель применяют в случае, если происходят существенные сезонные изменения.

Н

Неидентифицируемая модель – разновидность структурной модели системы одновременных уравнений, в которой структурные коэффициенты невозможно найти по приведенным коэффициентам.

Нелинейная связь – связь, выраженная уравнением криволинейной зависимости (гипербола, парабола и т.п.).

Нелинейная регрессия внутренне линейная – регрессия, которая с помощью соответствующих преобразований может быть приведена к линейному виду.

Нелинейная регрессия внутренне нелинейная – регрессия, которая не может быть сведена к линейной функции.

Несмещенность оценки – означает, что математическое ожидание остатков равно нулю.

Неполной (частичной) корректировки модель – разновидность динамической эконометрической модели, в которой учитывается ожидаемое значение результативного признака y_{t+1}.

О

Обратная связь – соотношение, при котором с увеличением факторного признака результативный признак уменьшается, и наоборот: уменьшение факторного признака приводит к увеличению результативного.

Однофакторная (парная) зависимость – это связь между одним признаком-фактором и результативным признаком.

Основная тенденция (тренд) – достаточно плавное и устойчивое изменение уровня явления во времени, более или менее свободное от случайных колебаний. Основную тенденцию можно представить либо аналитически, в виде уравнения (модели) тренда, либо графически.

Ошибка наблюдения – расхождение между фактическим и зарегистрированным значениями показателя в процессе статистического наблюдения.

П

Параболическая модель третьего порядка – модель вида

$$y = a + bx + cx^2 + dx^3 + \varepsilon.$$

Параметризация – определение вида экономической модели, выражение в математической форме взаимосвязи между ее переменными, формулирование исходных предпосылок и ограничений модели; оценка параметров построенной модели, делающих выбранную модель наиболее адекватной реальным данным.

Парная корреляция – это связь между двумя признаками (результативным и факторным или двумя факторными).

Парный коэффициент регрессии – показывает, на какую величину в среднем изменится результативный признак *y*, если переменную *x* увеличить на единицу измерения.

Парная регрессия – характеризует связь между двумя признаками: результативным и факторным.

Парный коэффициент детерминации – показывает, какая доля вариации переменной *y* учтена в модели и обусловлена влиянием на нее переменной *x*.

Поведенческие уравнения – описывают взаимодействие между экзогенными и эндогенными переменными в структурной форме системы одновременных уравнений.

Перспективная экстраполяция – предполагает продолжение ряда динамики на будущее, на основе выявления закономерности изменений уровней ряда в изучаемый период времени.

Предмет эконометрики – факты, формирующие развитие экономических процессов и явлений.

Производственная функция – характеризует связь между производственными факторами и величиной продукта.

Предопределенные переменные – лаговые и текущие экзогенные, а также лаговые эндогенные переменные модели.

Предпосылки МНК – математическое ожидание случайного отклонения равно нулю; дисперсия случайных отклонений постоянна; случайные отклонения являются независимыми друг от друга; случайное отклонение должно быть независимо от объясняющих переменных; модель является линейной относительно параметров.

Приведенная форма модели – один из способов записи системы одновременных уравнений, в котором каждая эндогенная переменная определена в виде линейной функции от всех предопределенных переменных.

Признак – основная отличительная черта, особенность изучаемого явления или процесса.

Причинно-следственные отношения – взаимосвязь социально-экономических явлений, когда изменение факторного признака оказывает воздействие на поведение результативного.

Проверка статистических гипотез о типе тренда – метод распознавания типа тренда, при котором проводится: сглаживание ряда уровней (скользящая средняя); вычисляют цепные абсолютные изменения $\Delta_i = Y_{i+1} - Y_i$ (для параболы – ускорения, для экспоненты – темпы роста); расчет по равным или примерно равным подпериодам средней величины того параметра, постоянство которого подтверждает выдвинутую гипотезу о типе тренда (средний абсолютный прирост – для прямой, среднее ускорение – для параболы, средний темп – для экспоненты); проверяется методом дисперсионного анализа или по t-критерию существенность различия средних значений параметра в разных подпериодах исходного ряда. Если различия средних признаются существенными, гипотеза о данном типе тренда отвергается и выдвигается следующая гипотеза в порядке усложнения: после отклонения прямой линии – об экспоненте; после отклонения экспоненты – о параболе; при отклонении параболы – о других типах линий.

Промежуточный мультипликатор – показатель модели авторегрессии, который определяет общее абсолютное изменение результата в момент времени (t + 1).

Пространственные данные – набор сведений по разным объектам, взятым за один и тот же период времени.

Прямолинейная зависимость – это связь, при которой с возрастанием величины факторного признака происходит равномерное возрастание (или убывание) величин результативного признака.

Р

Ранг – показатель, характеризующий порядковое место признака, расположенного в ранжированном ряду.

Ранжирование – процедура установления относительной значимости (предпочтительности) исследуемых объектов на основе их упорядочения.

Регрессионный анализ – способ установлений аналитической зависимости, в которой изменение среднего значения результативного признака обусловлено влиянием одного или нескольких независимых величин, а множество прочих факторов, также оказывающих влияние на результативный признак, принимается за постоянные или средние уровни.

Регулярная компонента – составляющая временного ряда, которая характеризует общую тенденцию ряда.

Результативный признак – зависимый признак, изменяющийся под влиянием связанного с ним факторного признака.

Ретроспективная экстраполяция – продолжение уровней временного ряда в прошлое.

Рыночные услуги – включают все услуги, реализуемые на рынке по экономически значимым ценам.

Ряд динамики (временные, хронологические ряды) – это ряд последовательно (в хронологическом порядке) расположенных статистических показателей, изменение которых имеет определенную тенденцию развития изучаемого явления; упорядоченные статистические данные по времени их получения. Он содержит лаговую составляющую.

Ряд Фурье – в гармониках Фурье исходным рядом является не первичный ряд за несколько лет, а усредненные значения месячных уровней, в которых исключены тренд и случайная компонента.

С

Сверхидентифицируемая модель – разновидность структурной модели системы одновременных уравнений, в которой структурные коэффициенты, выраженные через приведенные коэффициенты, имеют два и более числовых значений.

Связные временные ряды – временные ряды, показывающие зависимость результативного признака от одного или нескольких факторных.

Сезонная волна – это графическое изображение полученных индексов сезонности.

Сезонная компонента – компонента временного ряда, которая характеризует внутригодичные колебания показателя. В общем виде является циклической составляющей,

Сезонные колебания – колебания, периодически повторяющиеся в некоторое определенное время каждого года, месяца, дня или часа.

Система независимых уравнений – одна из разновидностей систем эконометрических уравнений, в которой каждый результативный признак является функцией одной и той же совокупности факторов; набор факторов в каждом уравнении системы может варьировать в зависимости от изучаемого явления.

Система одновременных уравнений – одна из разновидностей эконометрических моделей, состоящая из тождеств и регрессионных уравнений, в которых наряду с факторными признаками включены результативные признаки из других уравнений системы.

Система рекурсивных уравнений – одна из разновидностей систем эконометрических уравнений, в которой результативный признак одного уравнения системы в каждом последующем уравнении является фактором наряду с одной и той же совокупностью факторов.

Среднее квадратическое отклонение – корень квадратный из среднего квадрата отклонения (дисперсии).

Среднее линейное отклонение – средняя арифметическая величина из абсолютных значений отклонений, вариант признака от их среднего значения.

Средний абсолютный прирост – показатель, характеризующий среднюю абсолютную скорость роста (или снижения) уровня за отдельные периоды времени. Он показывает, на сколько единиц увеличился (или уменьшился) уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежемесячно и т. д.).

Средний темп прироста – относительный показатель, выраженный в процентах и показывающий, на сколько увеличился (или уменьшился) уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежемесячно и т.п.).

Средний темп роста – относительный показатель, выраженный в форме коэффициента и показывающий, во сколько раз увеличился уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежеквартально и т.п.).

Средняя хронологическая интервального ряда – исчисляется по формуле средней арифметической, причем при равных интервалах применяется средняя арифметическая простая, а при неравных – средняя арифметическая взвешенная.

Средняя хронологическая моментного ряда – исчисляется как сумма всех уровней ряда, поделенного на число членов ряда без одного, причем первый и последний члены ряда берутся в половинном размере.

Состоятельность оценок – характеризует увеличение их точности с увеличением объема выборки.

Спецификация – построение эконометрических моделей, т. е. представление экономических моделей в математической форме, удобной для проведения эмпирического анализа.

Статистическая зависимость – это связь, при которой каждому значению независимой переменной x соответствует множество значений зависимой переменной y, причем неизвестно заранее, какое именно значение примет y.

Стохастические модели – допускают наличие случайных воздействий на исследуемые показатели. Такой вид зависимости называется корреляционным.

Смыкание рядов – это объединение в один более длинный динамический ряд двух (или нескольких) рядов динамики, уровни которых исчислены по различной методологии или по различным границам территорий. Для смыкания необходимым условием является наличие за один период данных, рассчитанных по разной методологии (или в разных границах).

С распределенным лагом модель – содержит наряду с текущими значениями факторных переменных их лаговые значения.

Структурная форма модели – один из способов записи системы одновременных уравнений, который отражает реальный экономический объект или явление и показывает, как изменение любой экзогенной переменной определяет значения эндогенной переменной модели.

Т

Тенденция автокорреляции – вид тенденции временного ряда, который характеризует связь между отдельными уровнями ряда динамики; характеризует изменения связи между отдельными уровнями ряда динамики.

Тенденция дисперсии – представляет собой тенденцию изменения отклонений между эмпирическими уровнями и детерминированной компонентой ряда; вид тенденции временного ряда, который характеризует направление изменения отклонений между эмпирическими уровнями и детерминированной компонентой ряда.

Тенденция среднего уровня – вид тенденции временного ряда, который выражается обычно с помощью математического уравнения линии, вокруг которой варьируют фактические уровни исследуемого явления.

Темп прироста – относительный показатель динамики, отражающий, на сколько процентов уровень текущего периода больше или меньше уровня базисного или предыдущего периода.

Темп роста – относительный показатель, отражающий, во сколько раз уровень текущего периода больше или меньше уровня базисного или предыдущего периода, выраженный в коэффициентах или в процентах.

Тождество – одна из разновидностей структурных уравнений модели, которая устанавливает соотношение между эндогенными переменными; не содержит случайных составляющих и структурных коэффициентов.

Тренд – это основная, достаточно устойчивая тенденция во временном ряду, более или менее свободная от случайных колебаний; длительная тенденция изменения случайного процесса, определяющая основную тенденцию изменения экономических показателей.

У

Уровень ряда динамики — абсолютная (относительная, средняя) величина каждого члена динамического ряда.

Ф

Факторный признак — признак, оказывающий влияние на изменения результативного признака.

Функциональная зависимость — это связь, при которой каждому значению независимой переменной *x* соответствует точно определенное значение зависимой переменной *y*.

Ц

Циклические (или периодические) колебания — состоят в том, что значение изучаемого признака в течение какого-то времени возрастает, достигая определенного максимума, затем понижается, достигая определенного минимума, вновь возрастает до прежнего значения и т. д.

Цифровые метки (фиктивные переменные) — качественные переменные, преобразованные в количественные. Такого рода переменные в эконометрике принято называть фиктивными переменными.

Ч

Частный коэффициент корреляции — показывает степень тесноты связи между двумя признаками при фиксированном значении остальных факторных признаков.

Частная корреляция — это зависимость между результативным и одним факторным признаками или двумя факторными признаками при фиксированном значении других факторных признаков.

Частные показатели временного ряда — характеризуют явления изолированно, односторонне.

Частные уравнения регрессии — уравнения регрессии, которые связывают результативный признак с соответствующими факторами *x* при закреплении других учитываемых во множественной регрессии факторов на среднем уровне.

Э

Экзогенные (независимые) — это переменные, значения которых задаются извне модели; внешние по отношению к модели переменные, их значения определяются вне модели, и поэтому они считаются фиксированными, обозначаются обычно как *x*.

Эконометрика – это наука, предметом изучения которой является количественное выражение взаимосвязей экономических явлений и процессов.

Экономическая интерпретация модели – основные выводы и заключения на основе расчета и анализа частных коэффициентов эластичности, частных и множественного коэффициентов детерминации.

Эндогенные (зависимые) – это переменные, значения которых определяются внутри модели; переменные, значения которых определяются внутри модели и обозначаются обычно как y.

Эконометрическая модель – совокупность уравнений, описывающих связи между некоторыми экономическими показателями. Соотношения могут быть стохастическими (случайными) и детерминированными (зависящими от чего-либо).

Экстраполяция – метод научного исследования, заключающийся в распространении выводов, полученных из наблюдений над одной частью явления, на другую ее часть.

Эффективные оценки – таковыми считаются оценки, если они характеризуются наименьшей дисперсией.

ТЕСТЫ

1. Корреляционной называется связь, если:

а) каждому значению факторного признака соответствует вполне определенное неслучайное значение результативного признака;

б) каждому значению факторного признака соответствует множество значений результативного признака, т. е. определенное статистическое распределение;

в) каждому значению факторного признака соответствует целое распределение значений результативного признака;

г) каждому значению факторного признака соответствует строго определенное значение факторного признака.

2. Различают по аналитическому выражению связи:

а) обратные;

б) линейные;

в) криволинейные;

г) логистические.

3. Регрессионный анализ состоит в определении:

а) аналитической формы связи, в которой изменение результативного признака обусловлено влиянием одного или нескольких факторных признаков, а множество всех прочих факторов, также оказывающих влияние на результативный признак, принимается за постоянные и средние значения;

б) тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи);

в) статистической меры взаимодействия двух случайных переменных;

г) степени статистической связи между порядковыми переменными.

4. Частная корреляция — это:

а) зависимость результативного признака и двух и более факторных признаков, включенных в исследование;

б) связь между двумя признаками (результативным и факторным или двумя факторными);

в) зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков;

г) зависимость между качественными признаками.

5. Парный коэффициент корреляции не может принимать значение:

а) –0,973;

б) 0,005;

в) 1,111;

г) 0,721?

6. Связь между признаками У и Х можно считать тесной (сильной) при значении линейного коэффициента корреляции, равном:

а) –0,975;

б) 0,657;

в) –0,111;

г) 0,421?

7. Для оценки значимости коэффициента корреляции используют:

а) F-критерий Фишера;

б) t-критерий Стьюдента;

в) критерий Пирсона;

г) δ-критерий Дарбина–Уотсона?

8. Если парный коэффициент корреляции между признаками Y и X равен –1, то это означает:

а) отсутствие связи;

б) наличие обратной корреляционной связи;

в) наличие обратной функциональной связи;

г) наличие прямой функциональной связи?

9. Парный коэффициент корреляции между признаками Y и X принимает значение 0,675, соответственно детерминации равен:

а) 0,822;

б) –0,675;

в) 0,576;

г) 0,456?

10. В соответствии с методом наименьших квадратов минимизируется следующее выражение:

а) $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$;

б) $\sum_{i=1}^{n}(y_i - \hat{y}_i)$;

в) $\sum_{i=1}^{n}(y_i - \bar{y})^2$;

г) $\sum_{i=1}^{n}|y_i - \hat{y}_i|$.

11. В соответствии с МНК оценки параметров регрессии должны быть:

а) несмещенными;
б) гетероскедастичными;
в) эффективными;
г) состоятельными?

12. В уравнении линейной парной регрессии $\hat{y} = a + bx + \varepsilon$ параметр *b* означает:

а) усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов;

б) среднее изменение результативного признака при изменении факторного признака на 1%;

в) на какую величину в среднем изменится результативный признак у, если переменную х увеличить на единицу измерения;

г) какая доля вариации результативного признака у учтена в модели и обусловлена влиянием на нее переменной х?

13. В уравнении парной линейной регрессии значение параметра а определяется по формуле:

а) $\bar{y}\text{-}a\bar{x}$;

б) $\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$;

в) $\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$;

г) a_0x^n.

14. На сколько единиц своего измерения в среднем изменится у при увеличении *x* на одну единицу своего измерения, если уравнение регрессии имеет вид *y = 2,02+0,58x*:

а) увеличится на 2,02;

б) увеличится на 0,58;

в) увеличится на 2,80;

г) не изменится?

15. Для оценки значимости уравнения регрессии используют:

а) F-критерий Фишера;

б) t-критерий Стьюдента;

в) критерий Пирсона;

г) d-критерий Дарбина–Уотсона?

16. Множественный коэффициент корреляции изменяется в пределах:

а) $0 \le R_{yx_1x_2} \le \infty$;

б) $0 \le R_{yx_1x_2} \le 1$;

в) $-1 \le R_{yx_1x_2} \le 1$.

17. Множественный коэффициент детерминации изменяется в пределах:

а) $0 \le R^2_{yx_1x_2} \le 1$;

б) $1 \le R^2_{yx_1x_2} \le \infty$;

в) $-1 \le R^2_{yx_1x_2} \le 1$.

18. Что оценивает частный коэффициент корреляции:

а) тесноту связи между двумя переменными;

б) тесноту связи между тремя переменными;

в) тесноту связи между двумя переменными при фиксированном значении остальных факторов.

19. Коэффициент, указывающий в среднем процент изменения результативного показателя *y* при увеличении аргумента *x* на 1%, это:

а) коэффициент детерминации;

б) коэффициент регрессии;

в) коэффициент эластичности;

г) бета-коэффициент?

20. Если множественный линейный коэффициент корреляции $R_{yx1}x_2$ равен 0,75, то какой процент вариации зависимой переменной *y* учтен в модели и обусловлен влиянием факторов x_1, и x_2:

а) 56,2;

б) 75,0;

в) 37,5;

г) 62,4?

21. Укажите правильную характеристику параметра *k* тренда $y^t = a \cdot k^t$ **э тренда:**

а) постоянный цепной темп изменения уровней временного ряда;

б) среднее ускорение изменения анализируемого явления от периода (момента) к периоду (моменту) времени;

в) средний выровненный уровень ряда для периода (момента) времени, принятого за начало отсчета;

г) среднее изменение анализируемого явления от периода (момента) к периоду (моменту) времени.

22. Коэффициент a_2 параболического тренда $\hat{y}_i = a_o + a_1 t_i + a_2 t_i^2$ **характеризует:**

а) среднее изменение анализируемого явления от периода (момента) к периоду (моменту) времени;

б) средний выровненный уровень ряда для периода (момента) времени, принятого за начало отсчета;

в) постоянный цепной темп изменения уровней временного ряда;

г) среднее ускорение изменения анализируемого явления от периода (момента) к периоду (моменту) времени.

23. Коэффициент a_0 линейного тренда $\hat{y}_i = a_o + a_1 t_i + a_2 t_i^2$ **характеризует:**

а) средний выровненный уровень ряда для периода (момента) времени, принятого за начало отсчета;

б) среднее изменение анализируемого явления от периода (момента) к периоду (моменту) времени;

в) среднее ускорение изменения анализируемого явления от периода (момента) к периоду (моменту) времени;

г) постоянный цепной темп изменения уровней временного ряда.

24. Для получения устойчивой тенденции сезонных колебаний, на которых бы не отражались особенности развития процессов в конкретные периоды, индекс сезонности рассчитывают по формуле:

а) $\bar{I}_s = \dfrac{\sum I_s}{t}$;

б) $\bar{I}_s = \dfrac{y_i}{n}$;

в) $\bar{I}_s = \dfrac{\hat{y}}{\bar{y}}$;

г) $\bar{I}_s = \dfrac{\sum I_s}{y_s}$.

25. Укажите правильную функцию гиперболического тренда:

а) $\hat{y}_i = a_o + a_1 \cdot \dfrac{1}{t_i}$;

б) $\hat{y}_i = a_0 + a_1 t_i + a_2 t_i^2$;

в) $\hat{y}_i = \dfrac{1}{l^{a_0 + a_1 i} + 1}$;

г) $\hat{y}_i = y_{\min} + \dfrac{y_{\max} - y_{\min}}{l^{a_0 + a_1 i} + 1}$.

26. При моделировании временных рядов экономических показателей необходимо учитывать характер уровней исследуемых показателей:

1) аналитический;
2) конструкционный;
3) стохастический;
4) независящий от времени.

27. Состояние экономики в момент времени *t* описывается следующими характеристиками: C_t – ВВП, Y_t – уровень потребления, I_t – величина инвестиций, G_t– государственные расходы, T_t– величина налогов, R_t – реальная ставка процента. При этом величина инвестиций зависит от реальной ставки процента в предыдущем периоде, то есть в системе к предопределенным переменным системы относится лаговая экзогенная переменная, приведенное утверждение справедливо для модели…

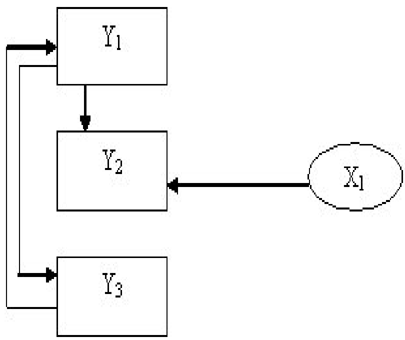
1. $$\begin{cases} C_t = a_0 + a_1 * (Y_t - T_t) + t \\ I_t = b_0 + b_1 * Y_t + b_2 * R_t \quad ; \\ Y_t = C_t + I_t + G_t \end{cases}$$

2. $$\begin{aligned} &C_t = a_0 + a_1 * (Y_t - T_t) + \xi_1 \\ &I_t = b_0 + b_1 * Y_{t-1} + b_2 * R_t + \xi_2; \\ &Y_t = C_t + I_t + G_t \end{aligned}$$

3. $$\begin{aligned} &C_t = a_0 + a_1 * (Y_t - T_t) + \xi_{t-1} \\ &I_t = b_0 + b_1 * Y_t + b_2 * R_t + \xi_{t-1}; \\ &Y_t = C_t + I_t + G_t \end{aligned}$$

4. $$\begin{aligned} &C_t = a_0 + a_1 * (Y_t - T_t) + \xi_1 \\ &I_t = b_0 + b_1 * Y_t + b_2 * R_{t-1} + \xi_2; \\ &Y_t = C_t + I_t + G_t \end{aligned}$$

28. Для указанной схемы взаимосвязей между переменными справедливы утверждения:



1) может быть описана с помощью системы рекурсивных уравнений;

2) включает 3 уравнения;

3) включает 4 уравнения;

4) может быть описана с помощью системы одновременных уравнений.

29. Эндогенные переменные:
1) могут быть объектом регулирования;
2) влияют на экзогенные переменные;
3) не зависят от экзогенных переменных;
4) могут коррелировать с ошибками регрессии.

30. Для оценки коэффициентов структурной формы модели не применяют метод наименьших квадратов:
1) обычный;
2) двухшаговый;
3) косвенный;
4) трехшаговый.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Кремер, Н. Ш., Путко, Б. А.* Эконометрика: Учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва: ЮНИТИ-ДАНА, 2002.

2. *Красе, М. С., Чупрынов, Б. П.* Математика для экономистов / М. С. Красе, Б. П. Чупрынов. – СПб.: Питер, 2004.

3. *Айвазян, С. А.* Эконометрика / С. А. Айвазян. – Москва: ЮНИТИ, 2001.

4. Эконометрика: учебник для вузов / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005.

5. Практикум для эконометрики / под ред. И. И. Елисеевой. – Москва: Финансы и статистика, 2005.

6. *Афанасьев, В. Н., Юзбашев, М. И.* Анализ временных рядов и прогнозирование: Учебник / В. Н. Афанасьев, М. И. Юзбашев. – М.: Финансы и статистика, 2001.

7. *Бородин, С. А.* Эконометрика: Учеб. пособие / С. А. Бородин. – Мн.: Новое знание, 2001.

8. *Доугерти, К.* Введение в эконометрику: Учебник. – 2-е изд. / К. Доугерти / пер. с англ. – М.: ИНФРА-М, 2004.

9. *Кобелев, Н. В.* Практика применения экономико-математических методов и моделей: Учеб.-практ. пособие / Н. В. Кобелев. – М.: ЗАО «Финстатинформ», 2000.

10. Статистика: учебник / под ред. И. И. Елисеевой. – М.: Высшее образование, 2006.

11. Статистика: учебник /под ред. И. И. Елисеевой. – М.: КНОРУС, 2006.

12. Статистика. Учебник /под ред. Р. А. Шмойловой. – М.: Финансы и статистика, 2004.

13. *Салин, В. Н., Чурилова, Э. Ю.* Курс теории статистики для подготовки специалистов финансово-экономического профиля: учебник / В. Н. Салин, Э. Ю. Чурилова. – М.: Финансы и статистика, 2007.

14. *Адамов, Е.* Экономика и статистка фирм: Учебник / под ред. С. Д. Ильенковой. – М.: Финансы и статистика, 2001.

15. *Балакина, Н. Н.* Статистика: Учеб.-метод. комплекс / Н. Н. Балакина. – Хабаровск: ИВЭСЭП, 2002.

16. *Дуброва, Т. А.* Статистические методы прогнозирования: Учеб. пособие / Т. А. Дуброва. – М.: ЮНИТИ-ДАНА, 2003.

17. *Кендам, М., Стьюартп, А.* Многомерный статистический анализ и временные ряды / М. Кендам, А. Стъюартп. – М.: Наука, 1976.

18. *Тихомиров, Н. П., Дорохина, Е. Ю.* Эконометрика: Учебник / Н. П. Тихомиров, Е. Ю. Дорохина. – М.: Экзамен, 2003.

Учебно-теоретическое издание

Кирин Игорь Григорьевич

ЭКОНОМЕТРИКА

Курс лекций
2-е издание исправленное и дополненное

Верстка Г. Сатлыкова
Корректор Л. Моисеева

Подписано в печать 22.06.12. Формат 60x84 $^{1}/_{16}$.
Бум. офсетная. Гарнитура «Newton». Печать цифровая.
Объем 3,80 уч.-изд. л. Тираж 100 экз. Заказ № 431.

Отпечатано в типографии ФГБОУ ВПО «ОГИМ»
460038, г. Оренбург, ул. Волгоградская, д. 16.
Тел./ факс: (3532) 30-50-00, доб. 127.