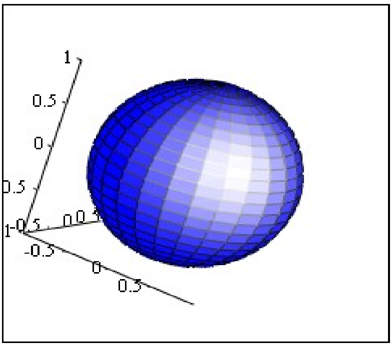
Курс лекций по

ЭКОНОМЕТРИКЕ



(X, Y, Z)

Жалалабат 2007

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
КЫРГЫЗСКОЙ РЕСПУБЛИКИ**

ЖАЛАЛАБАТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Курс лекций по

ЭКОНОМЕТРИКЕ

Жалалабат 2007

УДК 330
ББК 65
К 93

«Рекомендовано»
Методической комиссией
АЭФ

«Одобрено»
Методическим советом
ЖаГУ

Составители: преп. Мамыралиева А.Т.
канд.экон.наук, доц. Аскарова А.К.
ст. преп. Акжолова М.Ж.

Рецензент: д-р.физико-мат.наук, проф. Алыбаев С.К.

К 93 Курс лекций по эконометрике/Сост. А.Т. Мамыралиева и др.-Жалалабат: 2007.- 77 с.

ISBN 978-9967-09-151-1

ДАННЫЙ ЛЕКЦИОННЫЙ КУРС РАССМАТРИВАЮТ ПРОБЛЕМЫ ИЗУЧЕНИЯ И СОСТОЯНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ, СВЯЗАННЫЕ С ПОСТРОЕНИЕМ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ И ОПРЕДЕЛЕНИЯ ВОЗМОЖНОСТЕЙ ИХ ИСПОЛЬЗОВАНИЯ ДЛЯ ОПИСАНИЯ, АНАЛИЗА И ПРОГНОЗИРОВАНИЯ РЕАЛЬНЫХ ЭКОНОМИЧЕСКИХ ЗАДАЧ И ПРЕДНАЗНАЧАЕТСЯ СТУДЕНТАМ ПО СПЕЦИАЛЬНОСТИ «МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ», «МЕНЕДЖЕР - ОРГАНИЗАЦИИ»,

«ЭКОНОМИКА», А ТАКЖЕ ПРЕПОДАВАТЕЛЯМ ПО ЭТОЙ СПЕЦИАЛЬНОСТИ.

К 0601010000-07 УДК 330
ББК 65
ISBN 978-9967-09-151-1

ПРЕДИСЛОВИЕ

ЭКОНОМЕТРИКА ИССЛЕДУЕТ КОНКРЕТНЫЕ КОЛИЧЕСТВЕННЫЕ И КАЧЕСТВЕННЫЕ ВЗАИМОСВЯЗИ ЭКОНОМИЧЕСКИХ ОБЪЕКТОВ И ПРОЦЕССОВ С ПОМОЩЬЮ МАТЕМАТИЧЕСКИХ И СТАТИСТИЧЕСКИХ МЕТОДОВ И МОДЕЛЕЙ.

В ДАННОМ ЛЕКЦИОННОМ КУРСЕ РАССМАТРИВАЮТСЯ ПРОБЛЕМЫ ИЗУЧЕНИЯ И СОСТОЯНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ, СВЯЗАННЫЕ С ПОСТРОЕНИЕМ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ И ОПРЕДЕЛЕНИЯ ВОЗМОЖНОСТЕЙ ИХ ИСПОЛЬЗОВАНИЯ ДЛЯ ОПИСАНИЯ, АНАЛИЗА И ПРОГНОЗИРОВАНИЯ РЕАЛЬНЫХ ЭКОНОМИЧЕСКИХ ЗАДАЧ.

СОВРЕМЕННЫЕ ТРЕБОВАНИЯ К УРОВНЮ ЭКОНОМИЧЕСКОГО ОБРАЗОВАНИЯ ПРЕДПОЛАГАЮТ У ОБУЧАЕМЫХ НАЛИЧИЕ ШИРОКИХ МАТЕМАТИЧЕСКИХ И СТАТИСТИЧЕСКИХ ЗНАНИЙ, НЕОБХОДИМЫХ ПРИ РЕШЕНИИ ЗАДАЧ, СВЯЗАННЫХ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИРОВАНИЯ И КОЛИЧЕСТВЕННОГО АНАЛИЗА ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ.

ПЕРЕДОВЫЕ МАТЕМАТИЧЕСКИЕ И СТАТИСТИЧЕСКИЕ МЕТОДЫ ТАКЖЕ НЕРЕДКО ЯВЛЯЮТСЯ НЕОБХОДИМОЙ ПРЕДПОСЫЛКОЙ ДЛЯ ПРОВЕДЕНИЯ КАЧЕСТВЕННОГО АНАЛИЗА, ПРОГНОЗА И ОПРЕДЕЛЕНИЯ ОЦЕНОК СОСТОЯНИЯ И РАЗВИТИЯ ДЕЯТЕЛЬНОСТИ ЛЮБЫХ ПРЕДПРИЯТИЙ И ОРГАНИЗАЦИЙ.

МАТЕМАТИЧЕСКАЯ ПОДГОТОВКА ОБЫЧНО ОХВАТЫВАЕТ МАТЕМАТИЧЕСКИЙ АНАЛИЗ, МЕТОДЫ ЛИНЕЙНОЙ И НЕЛИНЕЙНОЙ ОПТИМИЗАЦИИ,

ИСПОЛЬЗОВАНИЕ ЛОГАРИФМОВ И ПОКАЗАТЕЛЬНЫХ ФУНКЦИЙ, УМЕНИЕ РАБОТАТЬ С ФОРМУЛАМИ, УРАВНЕНИЯМИ И ГРАФИКАМИ.

В СТАТИСТИЧЕСКУЮ ПОДГОТОВКУ, КАК ПРАВИЛО, НЕОБХОДИМО ВКЛЮЧАТЬ ЗНАКОМСТВО СО СТАТИСТИЧЕСКИМИ РАСПРЕДЕЛЕНИЯМИ, УМЕНИЕ ИЗМЕРЯТЬ ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ПАРАМЕТРЫ, - ТАКИЕ КАК СРЕДНЕЕ ЗНАЧЕНИЕ, СТАНДАРТНОЕ ОТКЛОНЕНИЕ И КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИЙ, ЗНАНИЕ МЕТОДОВ РЕГРЕССИОННОГО И КОРРЕЛЯЦИОННОГО АНАЛИЗА, МЕТОДОВ ДИСПЕРСИОННОГО, ФАКТОРНОГО И ДИСКРИМИНАНТНОГО АНАЛИЗА И УСЛОВИЯ ИХ ПРИМЕНЕНИЯ.

КУРС ЛЕКЦИЙ АДРЕСОВАН В ПЕРВУЮ ОЧЕРЕДЬ СТУДЕНТАМИ ДНЕВНЫХ ОТДЕЛЕНИЙ ЭКОНОМИЧЕСКИХ СПЕЦИАЛЬНОСТЕЙ. ОНИ НАЙДУТ НЕОБХОДИМЫЙ МАТЕРИАЛ ДЛЯ ИЗУЧЕНИЯ РАЗЛИЧНЫХ ВАРИАНТОВ ЭКОНОМЕТРИЧЕСКИХ КУРСОВ. СТУДЕНТЫ ЗАОЧНЫХ ОТДЕЛЕНИЙ, В ТОМ ЧИСЛЕ ПОЛУЧАЮЩИЕ ВТОРОЕ ОБРАЗОВАНИЕ ПО ЭКОНОМИКЕ И МЕНЕДЖМЕНТУ, СМОГУТ ИЗУЧИТЬ ОСНОВЫ ЭКОНОМЕТРИКУ И ПОЗНАКОМИТЬСЯ С ЕЕ КЛЮЧЕВЫМИ ИДЕЯМИ.

Введение

Современная экономическая теория, как на микро, так и на макро уровне, постоянно усложняющиеся экономические процессы привели к необходимости создания и совершенствования особых методов изучения и анализа. При этом широкое распространение получило использование моделирования и количественного анализа. На базе последних выделилось и сформировалось одно из направлений экономических исследований – эконометрика.

Эконометрика как наука расположена где–то между экономикой, статистикой и математикой, но ни одно из этих наук неспособна в отдельности, заменить эконометрику.

К основным задачам эконометрики можно отнести следующие:

- построение эконометрических моделей, то есть представление экономических моделей в математической форме. Данную проблему принято называть проблемой спецификации.

- оценка параметров построенной модели. Это этап параметризации.
- проверка качества найденных параметров модели и самой модели в целом. Иногда этот этап называют верификаций.
- использование построенных моделей для объяснения поведения исследуемых экономических показателей, прогнозирования и предсказания.

Эконометрические модели и методы сейчас – это не только мощный инструментарий для получения новых знаний в экономике, но и широко применяемый аппарат для принятия практических решений в прогнозировании, банковском деле, бизнесе. Развитие информационных технологий и специальных прикладных программ, совершенствование методов анализа сделали эконометрику мощнейшим инструментом экономических исследований.

Модели

Одно из определений эконометрики может звучать так: это наука, связанная с эмпирическим выводом экономических законов. То есть мы используем наблюдаемые значения или данные для того, чтобы получить количественные зависимости для экономических соотношений.

Но это – только малая часть задач, решаемых эконометрикой. Она также позволяет формулировать экономические модели, основываясь на эмпирических данных, оценивать неизвестные величины (параметры) в этих моделях, делать прогнозы и дает рекомендации по экономической политике.

В эконометрических исследованиях существенным является использование моделей. Модели должны быть “настолько простыми, насколько возможно, но не проще”, сказал Эйнштейн.

Рассмотрим пример.

Пусть необходимо проанализировать зависимость спроса Q на некоторый товар от цены P на этот товар. На основе экономической теории известно, что с ростом цены объем спроса сокращается. Опираясь на это утверждение можно предложить несколько математических зависимостей, отражающих этот факт. Например,

$Q = \alpha + \beta * P, \beta < 0$; $Q = \alpha * P^{\beta}, \beta < 0$;

$\ln Q = \alpha + \beta * \ln P, \beta < 0$;

Необходимо отметить, что любая из моделей будет лишь упрощением реальности и всегда содержит определенную погрешность. Поэтому из всех предлагаемых моделей с помощью статистических методов отбирается та, которая в наибольшей степени соответствует реальным эмпирическим данным и характеру зависимости.

Далее идет этап параметризации, то есть оценка параметров (в нашем случае α и β) так как эта оценка осуществляется на основе имеющихся статистических данных, то вопрос точности (качества) статистической информации является одним из ключевых при построении модели.

Затем проверяется качество найденных оценок, а также соответствие модели эмпирическим данным и теоретическим предпосылкам (этап верификации). Данный анализ в основном осуществляется по схеме проверки статистических гипотез. На этом этапе совершенствуется не только форма модели, но и уточняется состав ее объясняющих переменных (возможно спрос на товар определяется не только его ценой, но и другими факторами, например, располагаемым доходом).

Если модель удовлетворяет требованиям качества, то она может быть использована для прогнозирования, либо для анализа внутреннего механизма исследуемых процессов.

Математические модели широко применяются в бизнесе, экономике, общественных науках, исследовании экономической активности.

Математические модели позволяют более полно исследовать и понимать сущность происходящих процессов, анализировать их.

В эконометрических исследованиях используют разные типы моделей. Но можно выделить три основных класса моделей, которые применяются в эконометрике: модели временных рядов, регрессионные модели (с одним уравнением) и системы одновременных уравнений.

Типы данных

В эконометрических исследованиях используют два типа данных: пространственные данные (cross – sectional data) и временные ряды(time – seriesdata).

Пространственные данные – это данные, по какому- либо экономическому показателю, полученные для разных однотипных объектов (фирм, компаний, регионов).

Временные ряды – это данные, характеризующие один и тот же объект, но в различные моменты времени.

Примерами пространственных данных являются, например, объем производства, количество работников, прибыль и т.д. по разным предприятиям в один и тот же момент времени (пространственный срез).

Примерами временных рядов могут быть, например, ежеквартальные данные по прибыльности, объему выпускаемой продукции, средней заработной плате и т.д. для отдельного предприятия.

Тема 1. Предмет, задачи, критерии и принципы эконометрики

1.1. Предмет и задачи курса

На современном этапе экономического развития - деятельность в любой сфере (управлении, финансово-кредитной сфере, маркетинге, учете, аудите) требует от специалиста умения применить современные методы работы, знания достижений мировой экономической мысли, понимания научного языка. Большинство новых методов основано на эконометрических моделях, концепциях, приемах. Без глубоких знаний экономики научиться их использовать невозможно.

Специфической особенностью деятельности экономиста является работа в условиях неопределенности: недостатка информации и неполноты исходных данных. Для анализа такой информации требуются специальные методы, составляющие один из аспектов эконометрики. Центральной проблемой эконометрики является построение экономической модели и определение возможностей ее использования для описания, анализа и прогнозирования реальных экономических процессов.

Известный экономист Цви Гриллихес (1929-1999) писал: «Эконометрика является одновременно нашим телескопом и нашим микроскопом для изучения окружающего экономического мира».

Свидетельством всемирного признания эконометрики является присуждение Нобелевских премий за разработки в эконометрической области в 1969 г. Р.Фришу и Я.Тинбергену за разработку математических методов анализа экономических процессов; в 1980 – Л. Клейну за создание эконометрических моделей и их применение к анализу экономических колебаний и экономической политике; в 1989 г. – Т. Хаавельмо за прояснение вероятностных основ эконометрики и анализ одновременных экономических структур; в 2000 г.– Дж.

Хекману за развитие теории и методов анализа селективных выборок. Д. Макфаддену за развитие теории и методов анализа моделей дискретного выбора.

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

Термин «эконометрика» был впервые введен бухгалтером П. Цьемпой (Австро-Венгрия, 1910 г.) («эконометрия» - у Цьемпы).

Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика», выражая, таким образом, специфику, содержание эконометрики как науки: количественное выражение тех связей и соотношений, которые раскрыты и обоснованны экономической теорией.

Это новое научное направление возникло на стыке трех наук: экономической теории, статистики и математики. На дальнейшее развитие эконометрики большое воздействие оказывают новые информационные технологии.

Эконометрика входит в комплекс дисциплин «Экономико-математические методы». Ёе предметом является количественное выражение взаимосвязей и зависимостей экономических явлений и процессов, закономерностей экономики.

Первые попытки количественных исследований в экономике относятся к 17 веку. Однако эконометрика как наука сформировалась лишь в 20 – веке.

29 декабря 1930 г. по инициативе И. Фишера, Р.Фицера, Я. Тинбергена и др. на заседании Американской ассоциации развития науки было создано эконометрическое общество, на котором новой науке дали название – «эконометрика». В 1941 году появился первый учебник по эконометрике, который был создан Я.Тинбергеном.

1.2. Особенности эконометрического анализа

Становление и развитие эконометрического метода происходило на основе так называемой, высшей статистики – на методах парной и множественной регрессии, парной, частной и множественной корреляции, выделения тренда и других компонентов временного ряда.

В 30-е гг. XX века повсеместное увлечение множественной регрессией сменилось разочарованием. Строя уравнение множественной регрессии и, стремясь включить как можно больше объясняющих переменных, исследователи все чаще сталкивались с бессмысленными результатами. Причина заключалась в том, что изолированно взятое уравнение регрессии есть не что иное, как модель «черного ящика», поскольку в ней не раскрыт механизм зависимости выходной переменной Y от входных переменных X_i , а лишь констатируется факт наличия такой зависимости.

Для проведения правильного анализа нужно знать всю совокупность связей между переменными. Одним из первых подходов к решению этой задачи является конфлюэнтный анализ, разработанный в 1934 г. Р. Фришем. Он предложил изучать целую иерархию регрессий между всеми сочетаниями переменных.

На основе изменения коэффициентов регрессии b_i и множественного коэффициента детерминации R^2 он разделил все переменные на полезные, лишние и вредные. Переменная считалась полезной, если ее включение значительно повышало R^2; когда этого не происходило, она рассматривалась как лишняя. Если добавляемая переменная сильно изменяла коэффициенты регрессии b_i без заметного изменения R^2, то переменная относилась к вредным.

Эконометрика позволяет преодолеть искажающие воздействия ассимметричности, мультиколлинеарности, гетероскедастичности, автокорреляции, временных лагов и др. при исследовании экономических связей, зависимостей, закономерностей и тенденций.

Эконометрическое исследование включает решение следующих проблем:

- качественный анализ связей экономических переменных;
- выделение зависимых и независимых переменных;
- подбор данных;
- спецификация формы связи между показателями и факторами;
- оценка параметров модели;
- анализ мультиколлинеарности объясняющих переменных;
- введение фиктивных переменных;
- выявление автокорреляции, лагов;
- выявление тренда, циклической и случайной компонент;
- проверка остатков на гетероскедастичность;
- анализ структуры связей и построение системы одновременных уравнений;
- проверка условия идентификации;
- оценивание параметров системы одновременных уравнений;
- моделирование на основе системы временных рядов;
- построение реккурсивных моделей и т.д.

Эконометрическое моделирование включает следующие этапы:

а) постановка проблемы; б) получение данных, анализ их качества; в) спецификация модели; г) оценка параметров; д) интерпретация результатов.

1.3. Измерения в экономике

Понятие «эконометрика» включает экономические измерения. Признаками измерения состоит из: прежде всего, получение, сравнение и упорядочение информации. Это определение исходит из того, что измерение предполагает выделение некоторого свойства, по которому производится сравнение объектов в определенном отношении.

Другое понимание измерения исходит из числового выражения результата, т.е. измерение трактуется как операция, в результате которой получается численное значение величины.

Третий подход к измерению связан с обязательным наличием единицы измерения (эталона).

Любому измерению предшествует качественный анализ, учитывающий цели исследования качественный анализ необходим и после выполнения измерения.

Специфика экономических измерений состоит в наличии большого числа разнородных данных – разнородных ресурсов, разнородных результатов (например, товаров и услуг).

Нередко в экономических измерениях возникает задача отражения иерархии измерителей, которая выражается в выделении интегрального и частных показателей. Точность измерения – это его адекватность.

Для социально экономических измерений характерны специфические представления о точности. Экономику относят к неточным наукам, т.к. невозможно провести измерение с произвольно малой погрешностью.

По объективным причинам для социально-экономических измерений характерна низкая контролируемость их точности.

Основной базой данных для эконометрических исследований служат данные официальной статистики либо данные бухгалтерского учета. Т. о., проблемы экономического измерения – это проблемы статистики и учета.

ТЕМА 2. Введение в эконометрический анализ, основные его категории и понятия.

Любая экономическая деятельность несет в себе элемент стохастичности, что значит – осуществляя ту или иную экономическую операцию, анализируя динамику экономических показателей и т.д., ни один из специалистов не может быть уверен в конечном результате, поскольку все такие операции и показатели подвержены влиянию случайных факторов, то есть сами тоже являются случайными.

Причинами здесь являются – непредсказуемость доминирующего субъекта такой деятельности – человека и, в воздействии на любой экономической показатель огромного количества факторов. Многие из этих факторов человеком не контролируются.

Поэтому возникает проблема научного обоснования результатов экономической деятельности. Это можно осуществить лишь рассматривая экономические показатели и явления с учетом влияния на них случайных факторов, то есть, применяя аппарат теории вероятностей и математической статистики.

2.1. Вероятность. Случайная величина

Вероятность события A - это отношение числа m исходов, благоприятствующих появлению данного события, к общему числу n исходов, данного вероятностного эксперимента

$$P(A) = \frac{m}{n} \qquad (2.1)$$

Из определения вытекает очевидное неравенство $0 \leq P(A) \leq 1$

Любая деятельность в экономике по своей сути является вероятностной, то есть вероятностным экспериментом.

Событие – это любой исход, какого – либо вероятностного эксперимента.

Случайная величина (СВ) – это величина, которая может принимать то или иное значение, из некоторого множества значений.

Спрос на какую – либо продукцию, прибыль фирмы, объем экспорта за определенное время и т. д. являются случайными величинами.

Различают дискретные и непрерывные СВ. Дискретной называют такую СВ, которая принимает отдельные, изолированные значения с определенными вероятностями. Например, число покупателей в магазине в определенный момент времени, количество определенного товара, продаваемого ежедневно в магазине, число автомобилей на проспекте и т. д. является дискретными СВ. Непрерывной называется случайная величина, которая может принимать любые значения из некоторого конечного или бесконечного числового промежутка. Большинство СВ, рассматриваемых в экономике, имеют настолько большое число возможных значений, что их удобнее представлять в виде непрерывных СВ. Например, курсы валют, доход, объемы ВНП, ВВП и т. д.

Для описания дискретной СВ необходимо установить соответствие между всевозможными значениями СВ и их вероятностями. Такое соответствие называется законом распределения дискретной СВ. Его можно задать таблично, аналитически (в виде формулы) либо графически.

Например, табличное задание закона распределения дискретной СВ:

x	x_1	x_2	…	x_k
p_i	p_1	p_2	…	p_k

где $$\sum_{i=1}^{k} P_I = 1$$

Аналитически СВ задается либо функцией распределения, либо плотностью вероятностей.

Функцией распределения СВ X называется функция $F(x)$, которая определяется следующим образом:

$$F(x) = P(X < x),$$

то есть это есть вероятность того, что СВ X принимает значение меньшее, чем x.

Отметим некоторые свойства $F(x)$:

1. $0 \le F(x) \le 1$

2.$F(x)$ - неубывающая функция, то есть $(x_1 < x_2) \Rightarrow F(x_1) \le F(x_2)$

3. $\lim\limits_{x \to -\infty} F(x) = 0, \lim\limits_{x \to +\infty} F(x) = 1,$

4. Если СВ X принимает значения из отрезка $[a,b]$, то

$$F(x) = \begin{cases} 0, если \quad x \le a; \\ 1, если \quad x \ge b. \end{cases}$$

5. $P(a \le X \le b = F(b) - F(a)$

Плотностью вероятности (плотностью распределения вероятностей) непрерывной СВ X называют функцию

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x)}{\Delta x}$$

или из (5) свойства получаем

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x)$$

свойства плотности вероятности:

1. $f(x) \ge 0$

2. $P(a \le X \le b) = \int_a^b f(x)dx$

3. $F(x) = \int_{-\infty}^{x} f(t)dt$

4. $\int_{-\infty}^{+\infty} f(t)dt = 1$

2.2. Числовые характеристики случайных величин

Для удобства пользования СВ иногда удобнее бывает использовать их числовые характеристики. Важнейшими из них являются: математическое ожидание, дисперсия, среднее квадратическое отклонение. Математическое ожидание М(X) определяется следующим образом:

для дискретной СВ

$$M(X) = \sum_{i=1} x_i p_i, \quad (2.2)$$

для непрерывной СВ

$$M(x) = \int_{-\infty}^{+\infty} xf(x)dx \qquad (2.3)$$

Математическое ожидание характеризует среднее ожидаемое значение СВ. Однако для анализа СВ знания лишь среднего значения явно недостаточно. Существуют отличные друг от друга СВ, имеющие одинаковые математические ожидания.

Следовательно, нужна числовая характеристика, которая оценивает разброс возможных значений СВ относительно ее среднего значения (математического ожидания). Такой характеристикой является дисперсия.

Дисперсией D(*X*) СВ *X* называется математическое ожидание квадрата отклонения СВ от ее математического ожидания:

$$D(x)=M((x-M(x))^2)=M(x^2)-M^2(x) \quad (2.4)$$

При этом для дискретной СВ имеем:

$$D(x)=\sum_{i=1}^{k}(x_i-M(x))^2 * p_i=\sum_{i=1}^{k}x_i^2 p_i-M(x)^2 \qquad (2.5)$$

Для непрерывный СВ

$$D(x)=\int_{-\infty}^{+\infty}(x-M(x))^2 f(x)dx=\int_{-\infty}^{+\infty}x^2 f(x)dx-M^2(x) \qquad (2.6)$$

Так как дисперсия имеет размерность, равную квадрату размерности СВ, то вводится другая числовая характеристика-среднее квадратическое отклонение.

Средним квадратическим отклонением σ(x) СВ *X* называют величину:

$$\sigma(x)=\sqrt{D(x)} \qquad (2.7)$$

Для оценки разброса значений СВ в процентах относительно ее среднего значения, вводится коэффициент вариации *V*(*x*):

$$V(x)=\frac{\sigma(x)}{|M(x)|}100\% \qquad (2.8)$$

Меры разброса (дисперсия, среднее квадратическое отклонение, коэффициент вариации) кроме оценивания рассеивания значений СВ обычно применяются при изучении риска различных действий со случайным исходом: в финансовом анализе при оценивании различных активов и портфеля активов, при анализе риска инвестирования.

2.3. Законы распределений случайных величин

Зная конкретный закон распределения СВ (случайных величин) можно предвидеть вероятности попадания исследуемой СВ в определенные интервалы. Законов распределения много. Мы ограничимся рассмотрением тех, которые наиболее активно используются в эконометрическом анализе. К их числу относятся: нормальное распределение, распределение χ^2 , Стьюдента, Фишера.

Нормальное распределение

Нормальное распределение (распределение Гаусса) является предельным случаем почти всех реальных распределений вероятности.

Говорят, что СВ X имеет нормальное распределение, если ее плотность вероятности имеет вид:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (2.9)$$

Откуда получаем, что

$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt \qquad (2.10)$$

Как видно из формул (2.9) и (2.10) нормальное распределение зависит от параметров m и σ. При этом

$$m = M(x), \sigma = \sigma(x), D(x) = \sigma^2$$

Если СВ X имеет нормальное распределение с параметрами m и σ, то символически это записывается так:

$X \sim N(m,\sigma)$ или $X \sim N(m, \sigma^2)$

В случае, когда $m=0$ и $\sigma=1$ говорят о стандартном нормальном распределении.

Распределение χ^2 (хи – квадрат)

Пусть x_i $(i=1,2,\dots,n)$ -независимые нормально распредельные СВ с математическими ожиданиями m_i и среднеквадратическими отклонениями σ_i, соответственно, то есть $x_i \sim N(m_i,\sigma_i)$.

Тогда СВ $U_i = \dfrac{(x_i - m_i)}{\sigma_i}, (i = 1,2,\dots,n)$ являются независимыми СВ, имеющими стандартное нормальное распределение, $U_i \sim N(0,1)$.

Случайная величина χ^2 имеет хи–квадрат распределение с n-степенями свободы ($\chi^2 \sim \chi^2_n$), если

$$\chi^2 = \sum_{i=0}^{n} U_i^2 = U^2{}_1 + U^2{}_2 + \dots + U^2{}_n \ . \qquad (2.11)$$

(Число степеней свободы СВ определяется числом СВ, ее составляющих , уменьшенным на число линейных связей между ними).

Распределение χ^2 определяется одним параметром - числом степеней свободы ϑ.

$$M(\chi^2)=\vartheta;\ D(\chi^2)=2\vartheta.$$

Распределение Стьюдента

Пусть СВ $U \sim N(0,1)$, СВ V – независимая от U величина, распределенная по закону χ^2 с n-степенями свободы. Тогда величина

$$T = \frac{U}{\sqrt{V/n}} \quad (2.12)$$

имеет распределение Стьюдента (t- распределение) с n-степенями свободы.

$$M(T)=0; \quad D(T) = \frac{n}{n-2}$$

Распределение Стьюдента определяется одним параметром n. При $n>30$ распределение Стьюдента практически можно заменить нормальным распределением.

Распределение Фишера

Пусть V и W – независимые случайные величины, распределенные по закону χ^2 со степенями свободы $v_1= m$ и $v_2=n$ соответственно. Тогда величина

$$F = \frac{V/m}{W/n} \quad (2.13)$$

имеет распределение Фишера со степенями свободы $v_1= m$ и $v_2=n$. Т.о., распределение Фишера F определяется двумя параметрами – числами степеней свободы m и n:

$$M(F) = \frac{n}{(n-2)}(n-2),$$

$$D(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, (n>4).$$

При исследовании реальных экономических процессов приходится обрабатывать большие объемы статистических данных, которые по своей сути является СВ. Количество реализаций СВ ограничено, что не позволяет применять напрямую теоретические методы анализа. Поэтому в первую очередь используются методы и модели математической статистики, позволяющие получить необходимые знания, об исследуемом объекте.

2.4. Генеральная совокупность и выборка

Предположим, что изучается совокупность однородных объектов. Например, доход населения, количество покупателей в магазине в течение дня, данные о результатах голосования населения по какому – либо вопросу и т.д.

Генеральная совокупность - это всевозможные наблюдения интересующего нас показателя, все исходы случайного испытания.

Выборка – это часть генеральной совокупности, отобранная для изучения.

Число элементов совокупности называется ее объемом.

Изучение всей генеральной совокупности во многих случаях либо невозможно, либо нецелесообразно. Для анализа генеральной совокупности чаще всего используется выборка ограниченного объема.

Информация о генеральной совокупности, полученная на основании выборочного наблюдения, обычно обладает некоторой погрешностью. Это определяет две проблемы, связанные с выборками:

- организации выборочного наблюдения, чтобы полученная информация достаточно полно отражала пропорции генеральной совокупности (проблема репрезентативности выборки);
- использования результатов выборки для суждения по ним с наибольшей надежностью о свойствах и параметрах генеральной совокупности (проблема оценки).

Для обеспечения репрезентативности выборки применяют следующие способы отбора: простой отбор (последовательно отбирается первый, случайно попавшийся объект), типический отбор (объекты отбираются пропорционально представительству различных типов объектов в генеральной совокупности), случайный отбор, - например, с помощью таблицы случайных чисел и т. д.

2.5. Вычисление выборочных характеристик

При анализе конкретного показателя X в фиксированный момент времени наблюдаемые значения $x_1, x_2, \ldots, x_n$ обычно упорядочивают по неубыванию:

$x_1 \le x_2 \le \ldots \le x_n$ Разность между максимальным и минимальным значениями СВ X называется размахом выборки.

Если значение x_i встречается в выборке n_i раз, то число n_i называется частотой значения x_i, а величина $W_i = \dfrac{n_i}{n}$ - относительной частотой значения x_i.

Пусть объем генеральной совокупности равен N. Тогда величина $\bar{x}_R = \dfrac{1}{N}\sum_{i=1}^{N} x_i$ является генеральной средней. Генеральной дисперсией является величина $D_r = \dfrac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_r)^2$

Генеральным средним квадратическим отклонением является величина $\sigma_r = \sqrt{D_r}$. Так как в реальности чаще всего приходится работать с выборками, то приходится находить выборочные характеристики:

выборочное среднее:

$$\bar{x}_b = \frac{1}{n}\sum_{i=1}^{N} x_i .$$

выборочная дисперсия:

$$D_b = \frac{1}{n}\sum_{i=1}^{N}(x_i - \bar{x}_b)^2 .$$

выборочное среднее квадратическое отклонение:

$$\sigma_d = \sqrt{D_b} = \sqrt{\overline{x^2} - (\bar{x})^2} .$$

выборочный коэффициент вариации V: $V = \dfrac{\sigma_b}{x_b} 100\%$

2.6. Статистические выводы: оценки и проверка гипотез

Статистические выводы - это заключения о генеральной совокупности на основе выборки, случайно отобранной из генеральной совокупности. Например, анализируется такой показатель как доход (X) населения некоторого достаточно большого города. Этот анализ может быть осуществлен на основе выборки определенного объема (пусть n=1000). Для выборочных данных определяем средний доход $\bar{x} = \dfrac{\sum x_i}{n}$ и разброс $s^2 = \sum (x_i - \bar{x})^2 / n$.

Далее возникает естественный вопрос: можно ли ожидать, что аналогичные значения будут такими же для всего города? То есть можно ли обобщить результаты, полученные по выборке, на генеральную совокупность. В этом и суть статистических выводов.

На основе выборки можно получить лишь оценки параметров генеральной совокупности, так как оценки эти строятся на основе ограниченного набора данных. Естественно, значения оценок могут, изменяется от выборки к выборке. Процесс нахождения оценок по определенному правилу называется оцениванием.

Выделяют два типа оценивания: оценивание вида распределения и оценивание параметров распределения.

В качестве оценки вида распределения можно взять выборочное распределение, а в качестве оценок параметров распределения генеральной совокупности берутся их выборочные оценки.

Различают два вида оценок – точечные и интервальные.

После определения оценок обычно встает вопрос об их качестве и статистической значимости.

Пусть рассматривается генеральная совокупность наблюдаемой СВ X.

Для оценки ее параметра Θ из генеральной совокупности извлекается выборка объема n: $x_1, x_2, \ldots, x_n$. На основе этой выборки может быть найдена оценка Θ^* параметра Θ.

Точечной оценкой Θ^* параметра Θ называется числовое значение этого параметра, полученное по выборке объема n. Например, для нормального распределения параметрами являются математическое ожидание m и среднее квадратическое отклонение σ.

Оценками m и σ могут быть $m^* = \bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ и $\sigma^* = \sigma_\beta = \sqrt{\dfrac{1}{m}\sum_{i=1}^{n} (x_i - \bar{x})^2}$ соответственно.

Очевидно, что оценка Θ^* является функцией от выборки, то есть $\Theta^* = \Theta^*$ $(x_1, x_2, \ldots, x_n)$. А так как выборка носит случайный характер, то оценка Θ^* является СВ, принимающей различные значения для различных выборок . Любую оценку $\Theta^* = \Theta^*$ $(x_1, x_2, \ldots, x_n)$ называют статистической оценкой параметра Θ.

Качество оценок характеризуется следующими основными свойствами: **несмещенность, эффективность и состоятельность.**

Оценка Θ^* называется несмещенной оценкой параметра Θ, если ее математическое ожидание равно оцениваемому параметру: $M(\Theta^*) = \Theta$.

Оценка Θ^* называется эффективной оценкой параметра Θ, если ее дисперсия $D(\Theta^*)$ меньше дисперсии любой другой выборки объемом n.

Оценка параметра Θ называется асимптотически эффективной, если с увеличением объема выборки ее дисперсия стремится к нулю, то есть $D(\Theta^*_n) \to 0$ при $n \to \infty$.

Оценка Θ_n называется состоятельной оценкой параметра Θ, если Θ_n сходится по вероятности к Θ при $n \to \infty$, т.е. для любого $\varepsilon > 0$ при $n \to \infty$ $P(|\theta^*_x - \theta| < \varepsilon) \to 1$.

Иначе, состоятельной называется такая оценка, которая дает истинное значение при достаточно большом объеме выборки вне зависимости от значений входящих в нее конкретных наблюдений.

<u>Отметим некоторые свойства выборочных оценок.</u>

Доказано, что выборочное среднее $\bar{x}_\beta = \frac{1}{n}\sum_{i=1}^{n} x_i$ является несмещенной и состоятельной оценкой математического ожидания $M(X)$ генеральной совокупности.

Выборочная дисперсия $D_в = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}_\beta)^2$ является смещенной оценкой дисперсии $D(X) = \sigma^2$. Доказано, что $D_в = \sigma^2 \frac{(n-1)}{n}$ и это означает, что выборочная дисперсия оценивает генеральную дисперсию неточно.

Поэтому рекомендуется рассматривать исправленную дисперсию

$$S^2 = \frac{1}{n-1} D_в = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_\beta)^2$$

Исправленная дисперсия S^2 является несмещенной и состоятельной оценкой дисперсии D(X) СВ X.

Необходимо отметить, что при n >30 различие между $D_в$ и S^2 практически незначимо. Поэтому при большом объеме выборки оценки эти можно считать несмещенными.

2.7. Статистическая проверка гипотез

После построения, эконометрическая модель обычно требует многократного улучшения и уточнения. Для этого проводятся соответствующие расчеты по схеме статистической проверки гипотез.

Если закон распределения генеральной совокупности не известен, но есть основания предполагать, что он имеет определенный вид (назовем его A), то выдвигают гипотезу: генеральная совокупность, то есть СВ X, распределена по закону A. Например, можно выдвинуть предположение , что доход жителей некоторого города, региона, объем выпускаемой неким предприятием продукции имеют нормальный закон распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр Θ равен ожидаемому числу Θ_0, выдвигают гипотезу: $\Theta = \Theta_0$.

Статистической называют гипотезу о виде закона распределения или о параметрах известного распределения. В первом случае гипотеза называется непараметрической, а во второй - параметрической.

Гипотеза H_0 , подлежащая проверке, называется нулевой (основной). Наряду с ней рассматривают гипотезу H_1, которая будет приниматься, если отклоняется H_0. Такая гипотеза называется альтернативной. Если, например, проверяется гипотеза о равенстве параметра Θ значение Θ_0, то есть H_0: $\Theta = \Theta_0$, то в качестве альтернативной могут рассматриваться следующие гипотезы:

$$H_1^{(1)}: \Theta \neq \Theta_0;\, H_1^{(2)}: \Theta > \Theta_0;\, H_1^{(3)}: \Theta < \Theta_0;\, H_1^{(4)}: \Theta = \Theta_1; (\Theta_1 \neq \Theta_0)$$

Сущность проверки статистической гипотезы заключается в том, чтобы установить, согласуются или нет данные наблюдений и выдвинутая гипотеза.

Если при проверке гипотезы выборочные данные противоречат этой гипотезе H_0, то она отклоняются, в противном случае она не отклоняется. При этом возможны ошибки двух родов.

Ошибка первого рода состоит в том, что будет отвернута правильная нулевая гипотеза.

Ошибка второго рода состоит в том, что будет принята нулевая гипотеза, в то время как в действительности верна альтернативная гипотеза.

Исключить ошибки первого и второго рода невозможно в силу ограниченности выборки. Поэтому стремятся минимизировать потери от этих ошибок.

Вероятность совершить ошибку первого рода принято обозначать α (уровень значимости). Вероятность совершить ошибку второго рода обозначают, β. Тогда вероятность не совершить ошибку второго рода $(1-\beta)$ называется мощностью критерия.

Для проверки статистической гипотеза используют специально подобранную СВ (статистику, критерий), точное или приближенное значение которой известно .Эту величину обозначают:

U (или *Z*) – если она имеет стандартизированное нормальное распределение;

T - если она распределена по закону Стьюдента;

χ^2 - если она распределена по закону χ^2;

F – если она имеет распределение Фишера.

Для общности такую СВ будут обозначать через *K*.

Статистическим критерием называют СВ *K*, которая служит для проверки нулевой гипотезы. После выбора определенного критерия множество всех его возможных значений разбивают на два непересекающихся подмножества. Совокупность значений критерия, при которых нулевая гипотеза отклоняется, и другое - при которых она не отклоняется.

Первое подмножество называют критической областью, второе - областью принятия гипотезы.

Точки, разделяющие критическую область и область принятия гипотезы, называют критическими.

Тема 3. Корреляционный и регрессионный анализ – математический метод оценки взаимосвязей экономических явлений

3.1. Парная регрессия и корреляция в эконометрических исследованиях

3.1.1. Модель парной регрессии. Спецификация модели

Любой экономический показатель практически зависит от бесконечного количества факторов. Однако лишь ограниченное количество факторов действительно существенно воздействуют на исследуемый экономический показатель. Доля влияния остальных факторов столь незначительно, что их игнорирование не может привести к существенным отклонениям в поведении исследуемого объекта. Выделение и учет в модели лишь ограниченного числа реально доминирующих факторов является важной задачей качественного анализа, прогнозирования и управления экономической ситуаций.

Если в естественных науках большей частью имеют дело со строгими (функциональными) зависимостями, при которых каждому значению одной переменной соответствует единственное значение другой, то между экономическими переменными, в большинстве случаев, таких зависимостей нет. Поэтому в экономике имеют дело с корреляционными зависимостями.

В зависимости от количества факторов, включенных в уравнение регрессии, принято различать простую (парную) и множественную регрессии.

Простая регрессия представляет собой регрессию между двумя переменными y и x, т.е. модель вида

$$y = f(x),$$

где y – зависимая переменная (результативный признак); x – независимая, или объясняющая, переменная, (признак – фактор).

Строится простая (парная) регрессия в случае, когда среди факторов, влияющих на результативный показатель, есть явно доминирующий фактор.

Множественная регрессия соответственно представляет собой модель вида:

$y=f(x_1, x_2,\ldots,x_k)$ где x_i– признак – факторы.

Рассмотрим простейшую линейную модель парной регрессии:

$$y = a+bx+\varepsilon \quad (3.1)$$

Величина y, рассматриваемая как зависимая переменная, состоит из двух составляющих: неслучайной составляющей, $a+bx$ и случайного члена ε.

Случайная величина ε называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения.

Причин существования случайной составляющей несколько.

1. ***Не включение объясняющих переменных***. Соотношение между *y* и *x* является упрощением. В действительности существуют и другие факторы, влияющие на *y*, которые не учтены в (3.1). Влияние этих факторов приводит к тому, что наблюдаемые точки лежат вне прямой $y = a+bx$.

Часто встречаются факторы, которых следовало бы включить в регрессионное уравнение, но невозможно этого сделать в силу их количественной неизмеримости. Возможно, что существуют также и другие факторы, которые оказывают такое слабое влияние, что их в отдельности не целесообразно учитывать, а совокупное их влияние может быть уже существенным. Кроме того, могут быть факторы, которые являются существенными, но которые из-за отсутствия опыта таковыми не считаются. Совокупность всех этих составляющих и обозначено в (3.1) через ε.

2. ***Агрегирование переменных.*** Рассматриваемая зависимость (3.1) – это попытка объединить вместе некоторое число микроэкономических соотношений. Так как отдельные соотношения, имеют разные параметры, попытка объединить их является аппроксимацией. Наблюдаемое расхождение приписывается наличию случайного члена ε.

3. ***Выборочный характер исходных данных.*** Поскольку исследователи чаще всего имеет дело с выборочными данными при установлении связи между *y* и *x*, то возможны ошибки и в силу неоднородности данных в исходной статистической совокупности. Для получения хорошего результата обычно исключают из совокупности наблюдения с аномальными значениями исследуемых признаков. И в этом случае результаты регрессии представляют собой выборочные характеристики.

4. ***Неправильная функциональная спецификация.*** Функциональное соотношение между *y* и *x* математически может быть определено неправильно. Например, истинная зависимость может не

являться линейной, а быть более сложной. Следует стремиться избегать возникновения этой проблемы, используя подходящую математическую формулу, но любая формула является лишь приближением истинной связи y и x и существующее расхождение вносит вклад в остаточный член.

5. ***Возможные ошибки измерения.***

В парной регрессии выбор вида математической функции $y_x=f(x)$, может быть осуществлен графическим, аналитическим, экспериментальным методами.

Наиболее наглядным методом является графический. Он основан на поле корреляции.

Основные типы кривых, используемых при количественной оценке связей, представлены на рис. 3.1.

Кроме уже указанных используют также и другие типы кривых, например:

$$y = \frac{1}{a + bx};\ y = a + bx + \frac{c}{x};\ y = a + b\lg(x);\ y = \frac{1}{a + bx + cx^2}$$

Значительный интерес представляет аналитический метод выбора типа уравнения регрессии, который основан на изучении материальной природы связи исследуемых признаков.

Пусть, например, изучается потребность предприятия в электроэнергии y в зависимости от объема выпускаемой продукции x.

Общее потребление электроэнергии y можно подразделить на две части:

- не связанное с производством продукции a;
- непосредственно связанное с объемом выпускаемой продукции, пропорционально возрастающее с увеличением объема выпуска ($b{\cdot}x$).

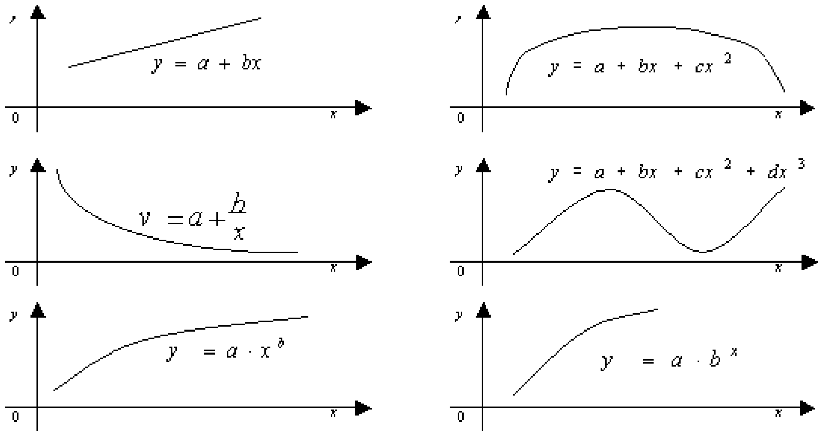


Рис 3.1. Основные типы кривых, используемые при количественной оценке связей между двумя переменными

Тогда зависимость потребления электроэнергии от объема продукции можно выразить уравнением регрессии вида.$y = a+bx$

Если разделим обе части уравнения на величину объема выпускаемой продукции (x), то получим выражение зависимости удельного расхода электроэнергии на единицу продукции ($z = y/x$) от объема выпущенной продукции (x) в виде уравнения гиперболы: z = b+a/x

При обработке информации на компьютере выбор вида уравнения регрессии осуществляется экспериментальным методом, т.е. путем сравнения величины остаточной дисперсии $D_{ост}$, рассчитанный при разных моделях.

В реальных условиях, как правило, всегда имеет место некоторое отклонение точек результативного признака относительно линии регрессии, обусловленное, присутствием случайного члена ε.

$$D_{ост} = \frac{1}{n}\sum (y - y_x)^2 \quad (3.2)$$

Поэтому для уравнения регрессии вычисляется величина суммы отклонений (y-y_x):

где y –фактические значения результативного признака,

y_x – расчетные значения, полученные по уравнению регрессии

Чем меньше величина $D_{ост}$, тем лучше уравнение регрессии описывает рассматриваемую корреляционную связь. Из разных математических функций выбирается та, для которой $D_{ост}$ является min.

В случае, когда $D_{ост}$ оказывается примерно одинаковой для нескольких функций, то предпочтение отдается более простым видам функций.

Обычно число наблюдений должно в 6-7 и более раз превышать число рассчитываемых параметров при переменной *x*.

3.1.2. Линейная регрессия сущность, оценка параметров

Линейная регрессия сводится к построению уравнения вида $y=a+b\cdot x$

Построение уравнения регрессии сводится в первую очередь к расчету его параметров - *a* и *b*. Они могут быть определены разными методами. Наиболее распространенным методом, является метод наименьших квадратов (МНК).

Допустим, что заданы *n* наблюдаемых значений результативного признака (*y*) и признака-фактора (x).

Следует отметить, что рассчитываются не истинные значения *a* и *b*, а только оценки, которые могут быть хорошими или плохими.

Возникает вопрос: существует ли способ достаточно точной оценки *a* и *b* алгебраическим путем?

Вначале на поле корреляции построим точки соответствующие наблюдаемым значениям *x* и *y* и прямую, выражающую линейную регрессию (рис.3.2).

$$\varepsilon_i = y_i - \hat{y}_{x_i}$$

Первым шагом является определение остатка для каждого наблюдения. Разность между фактическим и расчетным значением, соответствующим x_i, описывается как остаток в i-м приближении:

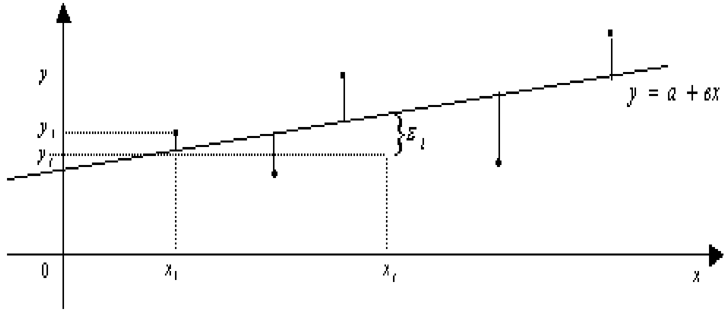


Рис.3.2 Точки рассеивания и прямая, выражающая линейную регрессию

Очевидно, что нужно построить такую линию регрессии, чтобы остатки были минимальными. Необходимо выбрать какой-то критерий подбора, который будет одновременно учитывать величину всех остатков.

Критерий минимизации суммы квадратов отклонений, фактических значений результативного признака (y) от расчетных (теоретических) $\hat{y}_к$:

$$\sum (y_i - \hat{y}_{x_i})^2 = \sum_i \varepsilon_i^2 \to \min \quad (3.3)$$

заложен в основу МНК. Обозначим $\sum \varepsilon_i^2$ через S, тогда

$$S = \sum_i (y_i - \hat{y}_{x_i})^2 = \sum (y - a - b \cdot x_i)^2 \quad (3.4)$$

Чтобы найти min (3.4), надо вычислить частные производные по каждому из параметров a и bи приравнять их к нулю:

$$1. \quad \begin{cases} \dfrac{dS}{da} = -2\sum y + 2na + 2b\sum x = 0 \\ \dfrac{dS}{db} = -2\sum yx + 2a\sum x + 2b\sum x^2 = 0 \end{cases} \quad (3.5)$$

Преобразуя систему (3.5), получаем следующую систему нормальных уравнений для оценки параметров a и b:

$$\begin{cases} n \cdot a + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum y \cdot x \end{cases} \quad (3.6)$$

Решая систему (3.6), получим

$$a = \bar{y} - b \cdot \bar{x},$$

$$b = \frac{\overline{x \cdot y} - \bar{y} \cdot \bar{x}}{\overline{x^2} - (\bar{x})^2}, \quad (3.7)$$

где $\bar{x} = \dfrac{\sum x}{n}$; $\bar{y} = \dfrac{\sum y}{n}$; $\overline{x^2} = \dfrac{\sum x^2}{n}$

Параметр b называется коэффициентом регрессии. Его величина показывает, насколько единиц изменится результат с изменением фактора на одну единицу.

Параметр a, вообще говоря, не имеет экономической интерпретации. Например, если a<0, то попытка его экономической интерпретации приводят к абсурду.

Зато можно интерпретировать знак при параметре a. Если, a>0, то относительное изменение результата происходит медленнее, чем изменение фактора.

3.1.3. Определение тесноты связи и оценка существенности уравнения регрессии

$$r = \frac{\overline{x \cdot y} - x \cdot y}{\sigma_x \cdot \sigma_y} = b \cdot \frac{\sigma_x}{\sigma_y} \quad (3.8)$$

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такового показателя выступает линейный коэффициент корреляции r

Одна из формул линейного коэффициента корреляции имеет вид:

где $\sigma_y = \sqrt{\bar{y}^2 - (\bar{y})^2}$; $\sigma_x = \sqrt{\bar{x}^2 - (\bar{x})^2}$; $\bar{y}^2 = \frac{\sum y^2}{n}$

Коэффициент корреляции находится в пределах: $-1 < r < 1$. Если b>0, то 0<r<1, и, наоборот, при b<0, -1< r<0.

Линейный коэффициент корреляции оценивает тесноту связи рассматриваемых признаков в ее линейной форме. Поэтому близость абсолютного значения линейного коэффициента корреляции к нулю еще не означает отсутствие связи между признаками. При нелинейном виде модели связь может оказаться достаточно тесной.

Квадрат линейного коэффициента корреляции называется коэффициентом детерминации. Он характеризует долю дисперсии результативного показателя y, объясняемую регрессией.

Соответственно величина $1-r^2$ характеризует долю дисперсии y, вызванную влиянием остальных, неучтенных в модели, факторов.

После того как построено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных ее параметров.

Оценка значимости уравнения регрессии в целом производится с помощью F-критерия Фишера.

С F-критерием тесно связана характеристика, называемая числом степеней свободы, которая применительно к исследуемой проблеме показывает, сколько независимых отклонений из n-возможных требуется для образования данной суммы квадратов.

Существует равенство между числом степеней свободы общей, факторной и остаточной суммы квадратов.

Число степеней свободы для факторной суммы квадратов равно 1, для общей суммы квадратов равно (n-1), для остаточной суммы квадратов составляет (n-2)

$$D_{общ} = \frac{\sum (y - \overline{y})^2}{n - 1};$$

$$D_{факт} = \frac{\sum (y_x - \overline{y})^2}{1};$$

$$D_{ост} = \frac{\sum (y - y_x)^2}{n - 2};$$

Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получаем дисперсию на одну степень свободы:

$$F = \frac{D_{фак}}{D_{ост}} \quad (3.9)$$

Сопоставляя факторную и остаточную дисперсию на одну степень свободы, получим величину F- отношения (F - критерий):

Величина F- критерия связана с коэффициентом детерминации r^2 :

$$F = \frac{r^2}{(1 - r^2)} \cdot (n - 2) \quad (3.10)$$

F - критерий для проверки нулевой гипотезы H_0:

$D_{факт} = D_{ост}$.

Т.е. если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Это дает основание считать, что влияние объясняющей переменной x модели несущественно, а, следовательно, общее качество модели невысоко.

Английским статистиком Снедекором разработаны таблицы критических значений F – отношений при разных уровнях существенности нулевой гипотезы и различном числе степеней свободы. Табличное значение F – критерия – это максимальная величина отношения дисперсии, которая может иметь место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы.

Если $F_{факт} > F_{табл}$, то нулевая гипотеза H_0 об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи.

Если $F_{факт} < F_{табл}$, то H_0 не отклоняется и уравнение регрессии считается статистически незначимым.

$$m_b = \sqrt{\frac{\sum (y-\hat{y}_x)^2}{(n-2)\cdot \sum (x-\bar{x})^2}}; \qquad (2.11)$$

$$m_a = \sqrt{\frac{\sum \left(y-\hat{y}_x\right)^2 \cdot \sum x^2}{(n-2)\cdot n \cdot \sum (x-\bar{x})^2}}. \qquad (2.12)$$

В линейной регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров. Для этого по каждому из параметров определяется его стандартная ошибка: m_b и m_a:

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 заданном уровне значимости α и числе степеней свободы (n-2)

Имеет место равенство: $t_b = \sqrt{F}$.

Для оценивания существенности параметра a определяется $t_a = \dfrac{a}{m_a}$ и его величина сравнивается с табличным значением.

Если табличное значение *t* – критерия превышает фактическое, то делается вывод о несущественности данного коэффициента, а если наоборот, табличное значение меньше фактического - вывод о существенности данного коэффициента.

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

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

Фактическое значение *t* – критерия Стьюдента определяется как

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

Т.о. проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

3.2. Нелинейная регрессия в экономике и ее линеаризация

3.2.1. Виды нелинейных регрессионных моделей, расчет их параметров

Хотя во многих практических случаях моделирование экономических зависимостей линейными уравнениями дает вполне удовлетворительный результат, однако ограничиться рассмотрением лишь линейных регрессионных моделей невозможно. Так близость линейного коэффициента корреляции к нулю еще не значит, что связь между соответствующими экономическими переменными отсутствует. При слабой линейной связи может быть очень тесной, например, не линейная связь. Поэтому необходимо рассмотреть и нелинейные регрессии, построение и анализ которых имеют свою специфику.

В случае, когда между экономическими явлениями существует нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных эконометрических моделей.

Различает две группы нелинейных регрессионных моделей:

модели, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам;

модели нелинейные по оцениваемым параметрам.

К первой группе относятся, например, следующие виды функций:

$y = a + bx + cx^2 + \varepsilon$ - полином 2-й степени;

$y = a + bx + cx^2 + dx^3 + \varepsilon$ - полином 3-й степени;

$y = a + \frac{b}{x} + \varepsilon$ - гипербола.

Ко второй группе относятся:

$y = a \cdot x^b \cdot \varepsilon$ - степенная;

$y = a \cdot b^x \cdot \varepsilon$ - показательная;

$y = e^{a+bx} \cdot \varepsilon$ - экспоненциальная и др. виды функций.

Классическим примером функций, относящихся к первой группе, являются кривые Филипса и Энгеля:

$y_x = a + \frac{b}{x}$ и $y_x = a - \frac{b}{x}$, соответственно.

Первая функция характеризует нелинейные соотношения между нормой безработицы *x* и процентом прироста заработной платы *y*. Из данной зависимости следует, что с ростом уровня безработицы темпы роста заработной платы в пределе стремится к нулю.

Вторая функция устанавливает закономерность – с ростом дохода доля расходов на продовольствие - уменьшается. Здесь *y,* обозначает - долю расходов на непродовольственные товары; *x* – доходы.

Первая группа нелинейных функций легко может быть линеаризована (приведены к линейному виду). Например, для полинома *к*-го порядка

$$y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \ldots + a_k \cdot x^k + \varepsilon$$

производя замену: $x = x_1, x^2 = x_2, x^3 = x_3, \ldots, x^k = x_k$ получим линейную модель вида

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_k \cdot x_k + \varepsilon$$

Аналогично могут быть линеаризованы и другие виды нелинейных функций 1-й группы, производя соответствующие замены.

Для оценки параметров нелинейных функций первой группы можно использовать, обычный МНК, аналогично, как и в случае линейных функций.

Иначе обстоит дело с группой регрессионных, нелинейных функций по оцениваемым параметрам. Данную группу функций можно разбить на две подгруппы:

нелинейные модели внутренне линейные;

нелинейные модели внутренне нелинейные.

Рассмотрим степенную функцию $y = a \cdot x^b \cdot \varepsilon$. Она нелинейна относительно параметров a и b. Однако ее можно считать внутренне линейной, так как, прологарифмировав ее можно привести к линейному виду:

$$\ln y = \ln a + b \ln x + \ln \varepsilon$$

Следовательно, ее параметры могут быть найдены обычным МНК.

Если модель представить в виде:

$y = a \cdot x^b + \varepsilon$, то модель становится внутренне нелинейной, т.к. ее невозможно преобразовать в линейный вид.

Внутренне нелинейной будет и модель вида: $y = a + bx^c + \varepsilon$

В эконометрических исследованиях, часто к нелинейным относят модели, только внутренне нелинейные по оцениваемым параметрам, а все другие модели, которые легко преобразуются в линейный вид, относятся к группе линейных моделей. Например, к линейным относят модель: $y = e^{a+bx} \cdot \varepsilon$, так как $\ln y = a + bx + \ln \varepsilon$

Если, модель внутренне нелинейна по параметрам, то для оценки параметров используются итеративные методы, успешность которых зависит от вида функции и особенностей применяемого итеративного подхода.

МНК в случае нелинейных функций, рассмотрим на примере оценки параметров степенной функции $y = a \cdot x^b \cdot \varepsilon$.

Прологарифмировав данную функцию, получим:

$\ln y = \ln a + b \ln x + \ln \varepsilon$ или, производя обозначения:

$y_1 = a_1 + bx_1 + \varepsilon_1$, где $y_1 = \ln y$; $a_1 = \ln a$; $x_1 = \ln x$; $\varepsilon_1 = \ln \varepsilon$

Применив МНК к полученному уравнению:

$$\begin{cases} \sum y_1 = n \cdot a_1 + b \sum x_1 \\ \sum y_1 \cdot x_1 = a_1 \sum x_1 + b \sum x_1^2, \end{cases} \quad \textit{или}$$

$$\begin{cases} \sum \ln y = n \cdot \ln a + b \sum \ln x \\ \sum \ln y \cdot \ln x = \ln a \sum \ln x + b \sum (\ln x)^2 \end{cases}$$

Параметр b определяется непосредственно из системы, а параметр, a – косвенным путем: $a = e^{a_1}$

3.2.2. Оценка корреляции для нелинейной регрессии

Оценка тесноты корреляционной зависимости в случае нелинейной регрессии производится с помощью индекса корреляции (R):

$$R = \sqrt{1 - \frac{\delta^2_{ост}}{\delta^2_y}},$$ где $$\delta^2_y = \frac{\sum (y - \bar{y})^2}{n},\quad \delta^2_{ост} = \frac{\sum (y - \hat{y}_x)^2}{n},$$

$\bar{y} = \frac{\sum y}{n}$, $\hat{y}_x$ – значения результативного признака, рассчитанные по уравнению регрессии.

Величина данного показателя находится в границах: $0 \le R \le 1$, чем она ближе к единице, тем теснее связь рассматриваемых признаков, тем надежнее найденное уравнение регрессии.

Величина R^2 называется индексом детерминации.

Оценка существенности индекса корреляции проводится, так же как и оценка надежности коэффициента корреляции. Индекс детерминации используется для проверки существенности в целом уравнения нелинейной регрессии по F-критерию Фишера:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m},$$ где R^2- индекс детерминации;

n - число наблюдений;

m - число параметров при переменных x.

Индекс детерминации R^2_{yx} можно сравнивать с коэффициентом детерминации r^2_{yx} для обоснования возможности применения линейной функции.

Возможность построения нелинейных моделей, как с помощью их приведения к линейному виду, так и путем использования нелинейной регрессии, значительно повышает универсальность регрессионного анализа, но и усложняет задачу исследователя.

Возникает вопрос: с чего начать - с линейной зависимости или с нелинейной, и если с последней, то, какого типа.

Если ограничиться парной регрессией, то можно построить график наблюдений y и x и принять решение. Однако очень часто несколько разных нелинейных функцией приблизительно соответствуют наблюдениям, если они лежать на некоторой кривой. А в случае множествен6ной регрессии невозможно даже построить график.

При рассмотрении альтернативных моделей с одним и тем же определением зависимой переменной процедура выбора достаточно проста. Наиболее разумным является оценивание регрессии на основе всех вероятных функций, и выбор функции, в наибольшей степени объясняющей изменения зависимой переменной. Если для одной модели коэффициент R^2 значительно больше, чем для другой, то вы сможете сделать оправданный выбор без особых раздумий, однако, если значения R^2 для двух моделей приблизительно равны, то проблема выбора существенно усложняется.

В этом случае следует использовать стандартную процедуру, известную под названием теста **Бокса – Кокса**.

Если необходимо сравнить модели с использованием y и lny в качестве зависимой переменной, то можно использовать вариант теста, разработанный Полом Зарембкой. Процедура включает следующие шаги:

1. Вычисляется среднее геометрическое значений y в выборке, (оно совпадает с экспонентой среднего арифметического lny):

$$y^* = \frac{y_1 \cdot y_2 \dots y_n}{n}$$

2. Пересчитываются наблюдения y, т.е. они делятся на это значение, то есть $y_i^* = \dfrac{y_i}{y^*}$

3. Оценивается регрессия для линейной модели с использованием y^*_i вместе y_i и для логарифмической модели с использованием $ln(y^*_i)$ вместо $ln(y_i)$. Теперь значения суммы квадратов отклонений для двух регрессий сравнимы, и, следовательно, модель с меньшей суммой квадратов отклонений обеспечивает лучшее соответствие.

3.3. Множественная регрессия и корреляция

3.3.1. Множественная регрессия. Отбор факторов при построении ее модели

На любой экономической показатель чаще всего оказывает влияние не один, а несколько факторов.

Парная (однофакторная) регрессия является частным случаем множественной регрессии. Схематически модель множественной регрессии записывается в виде: $y = f(x_1, x_2, \ldots x_n)$ где у результативный экономический показатель, $x_1, x_2, \ldots x_n$ - показатели факторы.

Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функций издержек производства, в макроэкономических расчетах и при решении других вопросов в различных экономических сферах. В настоящее время множественная регрессия – один из наиболее распространенных методов в эконометрике.

Основная цель множественной регрессии - построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели (выбор факторов, вида уравнения и др.)

Факторы, включаемые в модель множественной регрессии, должны отвечать следующим требованиям:

- должны быть количественно измеримы;
- не должны быть интеркоррелированы или находится в функциональной зависимости;
- в одну модель нельзя включать совокупный фактор и образующие его частные факторы, что может привести к неоправданному увеличенному их влияние на зависимый показатель, к искажению реальной действительности;
- количество включаемых в модель факторов не должно превышать одной трети числа наблюдений в выборке.

Отбор факторов обычно осуществляется в две стадии: на первой подбираются факторы, исходя из сущности проблемы; на второй – на основе матрицы показателей корреляции определяют t - 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 $r_{x_i y_j} \geq 0,7$.

Из двух явно коллинеарных факторов, уравнения регрессии - рекомендуется исключить один. Предпочтение при этом отдается тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.

Рассмотрим пример. Для зависимости $y = f(x_1, x_2, x_3)$ задана матрица парных коэффициентов корреляции:

	У	x_1	X_2	x_3
У	*1*			
x_1	*0,8*	*1*		
x_2	*0,7*	*0,8*	*1*	
x_3	*0,6*	*0,5*	*0,2*	*1*

Из таблицы, очевидно, что факторы x_1 и x_2 коррелированны друг с другом. В уравнение целесообразно включить фактор x_2, а не x_1, так как корреляция x_2 с y - слабее, чем корреляция фактора x_1 с y $(r_{yx_2} < r_{yx_1})$,но зато $r_{x1x3}>r_{x2x3}$. Поэтому в уравнение множественной регрессии включаются факторы x_2 и x_3.

При отборе влияющих факторов используются статистические методы отбора. Так, существенного сокращения числа влияющих факторов можно достичь с помощью пошаговых процедур отбора переменных. Ни одна их этих процедур не гарантирует получения оптимального набора переменных. Однако при практическом применении они позволяют получить достаточно хорошие наборы существенно влияющих факторов.

Наиболее широкое применение получили следующие методы отбора факторов: метод исключения, метод включения, шаговый регрессионный анализ.

Метод исключения предполагает построение уравнения, включающего всю совокупность переменных, с последующим последовательным (пошаговым) сокращением числа переменных в модели до тех пор, пока не выполнится некоторое, наперед заданное, условие. Суть метода включения состоит – в последовательном включении переменных в модель до тех пор, пока регрессионная модель не будет отвечать заранее установленному критерию качества. Последовательность включения определяется с помощью частных коэффициентов корреляции: переменные, имеющие относительно исследуемого показателя большие значения частного коэффициента корреляции, первыми включаются в регрессионное уравнение.

Шаговый регрессионный анализ состо1ит в исключении ранее введенного фактора. Матрица частных коэффициентов корреляции наиболее широко используется в процедуре отсева факторов.

Уравнения множественной регрессии как парной регрессии могут быть: линейными и нелинейными.

Наиболее часто используются линейная и степенная функции:

$\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon$ - линейная,

$\hat{y} = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot ... \cdot x_p^{b_p} + \varepsilon$ - степенная.

Линеаризуемы и широко применяются и следующие уравнения множественной регрессии:

$y = e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{p1} x_p + \varepsilon}$ - экспоненциального вида:

$y = \dfrac{1}{\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon}$ - гиперболического вида.

В качестве моделей множественной регрессии могут быть использованы и другие виды математических функций, которые могут быть преобразованы к линейному виду, например:

$$y = a + b_1 \cdot \sqrt{x_1} + b_2 \cdot \frac{1}{x_2^2} + b_3 \cdot \ln x_3 + \varepsilon$$

Обозначив $Z_1 = \sqrt{x_1}$; $Z_2 = \dfrac{1}{x_2^2}$; $Z_3 = \ln x_3$,

получаем линейное уравнение множественной регрессии

$$y = a + b_1 \cdot Z_1 + b_2 \cdot Z_2 + b_3 \cdot Z_3 + \varepsilon .$$

3.3.2. Расчет параметров и характеристик модели множественной регрессии

Параметры уравнения множественной регрессии оцениваются, как и в парной регрессии, МНК. При его применении строится система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии.

Для уравнения $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon$ система нормальных уравнений составит:

$$\begin{cases} \sum y = n \cdot a + b_1 \sum x_1 + b_2 \sum x_2 + ... + b_p \sum x_p \\ \sum y \cdot x_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + ... + b_p \sum x_p x_1; \\ .... \quad .... \quad .... \quad .... \quad .... \\ \sum y \cdot x_p = a \sum x_p + b_1 \sum x_p x_1 + b_2 \sum x_p x_2 + ... + b_p \sum x_p^2; \end{cases}$$

При нелинейной регрессии, приводимой к линейному виду, ее параметры также можно определить МНК с той лишь разницей, что он используется не к исходной информации, а к преобразованным данным.

Ценность эконометрических моделей состоит в том, что они позволяют не только выявить связи и зависимости, выразить их на языке математики, дать экономическое истолкование параметрам, но и в том, что позволяют рассчитать ряд характеристик.

Наиболее важными из них являются следующие:

- предельная эффективность показателя-фактора;
- коэффициент эластичности;
- изокванта;
- предельная норма заменяемости одного фактора другим;
- изоклинал;
- индексы корреляции и детерминации;
- стандартная ошибка и другие.

Рассмотрим сущность и методику расчета каждого из перечисленных характеристик.

Предельная эффективность показывает - на сколько абсолютных единиц измениться результативный показатель, если данный фактор увеличиться на одну абсолютную единицу, а остальные факторы останутся неизменными. Предельная эффективность представляет собой частную производную по показателю-фактору, т.е. $\frac{\partial Y}{\partial X_i}$, где i = 1,2,…,n.

Коэффициент эластичности показывает - на сколько процентов измениться результативный показатель, если данный показатель-фактор измениться на один процент, а остальные факторы останутся неизменными.

Формула для расчета коэффициента эластичности (Ex_i) имеет вид

$$Ex_i = \frac{\partial Y}{\partial X_i} \cdot \frac{X_i}{Y}$$

Следует обратить внимание на следующие частные случаи:

-в случае линейной зависимости предельная эффективность фактора равна коэффициенту регрессии, т.е. $\frac{\partial Y}{\partial X_i} = a_i \qquad i = 1,2,...,n;$

-в случае зависимости степенного вида коэффициент эластичности показателя-фактора равен коэффициенту регрессии, т.е.

$$Ex_i = b_{i_i}, \quad i=1,2,\dots,n .$$

Изокванта, предельная норма заменяемости одного фактора другим, изоклинал - характеристики, рассчитываемые только для многофакторных моделей.

Изоквантой называют множество сочетаний значений показателей-факторов, при которых результативный показатель принимает одно и тоже значение. Чтобы найти изокванту надо:

- принять Y за константу (Y= const);
- выразить один из факторов через остальные.

Для каждой эконометрической модели можно построить «семейство» изоквант.

Предельная норма заменяемости одного фактора другим позволяет-определить, сколько единиц одного фактора требуется для замены одной единицы другого фактора. Чтобы рассчитать предельную норму заменяемости надо:

- найти изокванту;
- определить частную производную одного фактора по другому, т.е. $\partial X_l/\partial X_k$, где $l \neq k$, l и $k \in i = 1,2,\dots,n$.

Изоклинал – это множество сочетаний значений показателей-факторов, при которых предельная норма заменяемости принимает одно и тоже значение. Чтобы найти изоклинал, надо:

- найти предельную норму заменяемости;

- принять предельную норму заменяемости за константу $\frac{\partial X_i}{\partial X_k} = const$

- выразить один из факторов через остальные.

3.3.3. Частные уравнения множественной регрессии. Индексы множественной и частной корреляции и их расчет

На основе линейного уравнения множественной регрессии $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon$ могут быть найдены частные уравнения регрессии:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, ..., x_p} = f(x_1) \\ y_{x_2 \cdot x_1, x_3, ..., x_p} = f(x_2) \\ ...... \quad ...... \quad .... \\ y_{x_p, x_2, x_2, ..., x_{p-1}} = f(x_p) \end{cases}$$

т.е. уравнения регрессии, которые связывают результативный признак с соответствующими факторами x_i при закреплении других учитываемых во множественной регрессии факторов на среднем уровне.

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – коэффициента детерминации. Показатель множественной корреляции характеризует тесноту совместного влияния факторов на результат.

Независимо от вида уравнения индекс множественной корреляции рассчитывается по формуле:

$$R_{yx_1x_2...x_p} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2_y}},$$

где σ^2_y - общая дисперсия результативного признака,

$\sigma^2_{ост}$ -статочная дисперсия для уравнения $У=f(x_1,x_2,... \ x_p)$

$0 \le R_{yx_1x_2...x_p} \le 1$

Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов.

Для расчета индекса множественной корреляции можно пользоваться и следующей формулой:

$$R^2_{yx_1x_2\ldots x_p} = 1 - \frac{\sum \left(y - \hat{y}_{yx_1x_2\ldots x_p}\right)^2}{\sum \left(y - \bar{y}\right)^2},$$

где y - фактические значения результативного показателя;

$\hat{y}_{yx_1,\ldots,x_p}$ - значения результативного показателя, рассчитанные по уравнению регрессии;

$\bar{y}$ - среднее арифметическое значение результативного показателя.

Сравнивая индексы множественной регрессии и парной корреляции, можно сделать вывод о целесообразности включения в уравнение регрессии того или иного фактора. В частности, если дополнительно включенные в уравнение множественной регрессии факторы третьестепенны, то индекс множественной корреляции практически совпадает с индексом парной корреляции.

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

Рассмотрим пример. Пусть зависимость объема продукции y от затрат труда x^1 задается уравнением: $\hat{y}_{x_1} = 2{,}75 + 3{,}5x, \qquad r_{yx_1} = 0{,}58$

Допустим, что дополнительный фактор x_2- техническая оснащенность производства – преобразовал уравнение к виду:

$$\hat{y}_{x_1x_2} = 20{,}2 + 2{,}8x_1 + 0{,}2x_2$$

Тогда остаточные дисперсии S^2 для этих уравнений определяются соответственно следующими формулами:

$$S^2_{yx_1} = \frac{\sum\left(y_i - \hat{y}_{x_1}\right)^2}{n}; \quad S^2_{yx_1x_2} = \frac{\sum\left(y_i - \hat{y}_{x_1x_2}\right)^2}{n}$$

Предположим, что. $S^2_{yx_1x_2} = 3{,}7$, $S^2_{yx_1} = 6$.

Уменьшение остаточной дисперсии за счет дополнительного включения фактора x_2 составит: $S^2_{yx_1} - S^2_{yx_1x_2} = 2{,}3$

Чем больше доля полученной разности в остаточной вариации, тем теснее связь между y и x_2, при неизменности действия фактора x_1

Величина, рассчитываемая формулой:

$$r_{yx_2x_1} = \sqrt{\frac{S^2_{yx_1} - S^2_{yx_1x_2}}{S^2_{yx_1}}}$$

называется индексом частной корреляции для фактора x_2:

Аналогично определяется индекс частной корреляции для фактора $x_{1.}$

$$r_{yx_1x_2} = \sqrt{\frac{S^2_{yx_2} - S^2_{yx_1x_2}}{S^2_{yx_2}}}$$

Если в нашем примере предположить, что $S^2_{yx2} = 5$, то частные коэффициенты корреляции составят: $r^2_{yx_1x_2} = 0{,}51$;. На их основе можно делать вывод: более сильное воздействие на объем продукции оказывает техническая оснащенность предприятий.

В общем случае при наличии p факторов формула для расчета индекса частной корреляции имеет вид:

$$r_{yx_ix_1x_2..x_{i-1}x_{i+1}...x_p} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2..x_p}}{1 - R^2_{yx_ix_1x_2..x_{i-1}x_{i+1}..x_p}}},$$

где $R^2_{yx_1x_2...x_p}$ - множественный коэффициент детерминации всего комплекса p факторов с результатом,

$R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_p}$ - тот же показатель детерминации, но без введения в модель фактора x_j.

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

Фишера: $F = \frac{D_{факт}}{D_{ост}} \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$

где R^2 - коэффициент (индекс) множественной детерминации;

m - число параметров при переменных x;

n - число наблюдений.

Если оценивается значимость влияния фактора x_i в уравнении регрессии, то определяется частный F- критерий:

$$F_{x_i} = \frac{R^2_{yx_1x_2\ldots x_p} - R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p}}{1 - R^2_{yx_1x_2\ldots x_p}} \frac{(n-m-1)}{1}.$$

Значимость коэффициентов чистой регрессии производится по t - критерию Стьюдента.

Если до сих пор в качестве факторов мы рассматривали только экономические переменные, принимающие количественные значения, то возможно, может оказаться необходимым включить в модель фактор, имеющий два или более качественных уровней. Например, такие атрибутивные признаки как профессия, пол, образование климатические условия и т.д. имеют несколько качественных уровня. Чтобы ввести такие переменные в модель необходимо их преобразовать в количественные переменные. Переменные такой конструкции называются фиктивными.

Рассмотренная модель с фиктивными переменными, выступающими как факторы, обладает наибольшими прогностическими возможностями. Однако на практике может возникнуть необходимость построения модели, в которой фиктивная переменная должна играть роль результата. Подобного рода модели применяются в социологии, при обработке данных социологических опросов. В качестве y - рассматриваются ответы на вопросы, данные в альтернативной форме: «да» или «нет», т.е. зависимая переменная y, имеет два значения *1* - («да») и *0* - («нет»).

3.3.4. Обобщённый метод наименьших квадратов. Гомоскедастичность и гетероскедастичность

При построении модели, например, линейного вида $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_p \cdot x_p + \varepsilon$ случайная величина ε представляет собой ненаблюдаемую величину. Для разных спецификаций модели разности $(y - \hat{y}_x)$ между теоретическими и фактическими значениями могут меняться. В задачу регрессионного анализа входит не только построение самой модели, но и исследование случайных отклонений ε_i т.е. остаточных величин. После построения уравнения регрессии проводится проверка наличия у оценок ε_i некоторых свойств. Эти свойства оценок, полученных МНК, имеют очень важное практическое значение в использовании результатов регрессии и корреляции.

Коэффициенты регрессии b_i, найденные на основе системы нормальных уравнений и представляющие собой выборочные оценки характеристики силы связи, должны обладать свойством несмещености. Несмещенность оценки означает, что математическое ожидание остатков равно нулю.

Это означает, что найденный параметр регрессии b_i, можно рассматривать как среднее значение возможных значений коэффициентов регрессии с несмещенными оценками остатков.

Для практических целей важны не только несмещенность, но и эффективность оценок. Оценки считаются эффективными, если они характеризуются наименьшей дисперсией.

Для того, чтобы доверительные интервалы параметров регрессии были реальными, необходимо, чтобы оценки были состоятельными. Состоятельность оценок характеризует увеличение их точности с увеличением объема выборки.

Исследования остатков ε_i предполагают проверку наличия следующих пяти предпосылок МНК:

случайный характер остатков;

нулевая средняя величина остатков, не зависящая от x_i;

гомоскедастичность–дисперсия каждого отклонения ε_i

одинакова для всех значений x;

отсутствие автокорреляции остатков.

Значения остатков ε_i распределены независимо друг от друга; остатки подчиняются нормальному распределению.

Если распределение случайных остатков ε_i не соответствует некоторым предпосылкам МНК, то следует корректировать модель.

Прежде всего, проверяется случайный характер остатков ε_i.

Строится график зависимости остатков ε_i от теоретических значений результативного признака (рис.3.3)

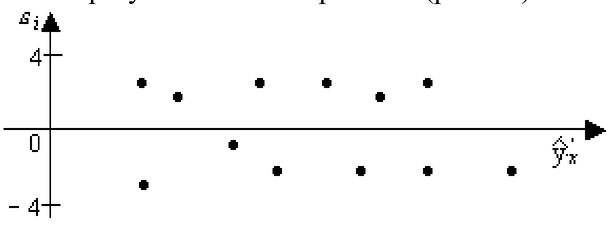


Рис 3.3.Зависимость случайных остатков ε_i.от теоретических значений y_x

Если на графике получена горизонтальная полоса распределения остатков, то остатки представляют собой случайные величины и МНК оправдан, теоретические значения y_x хорошо аппроксимируют фактические значения y.

Возможны следующие случаи: если ε_i. зависит от y_x то:

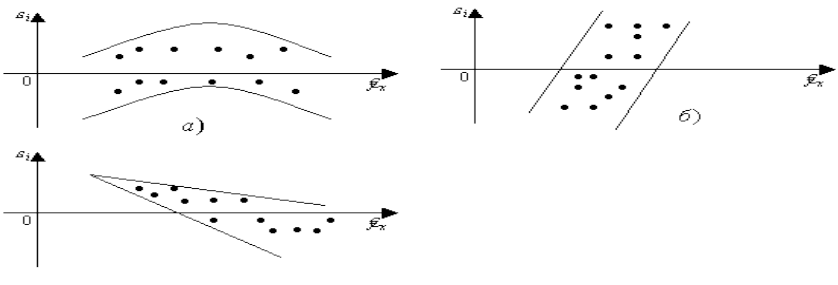
остатки ε_i. не случайны (рис.3.4а)

остатки ε_i. не имеют постоянной дисперсии (рис.3.4в)

остатки ε_i. носят систематический характер (рис. 3.4б).

В этих случаях *(а,б,в)* необходимо либо применить другую функцию, либо вводить дополнительную информацию и заново строить уравнение регрессии до тех пор, пока остатки ε_i не будут случайными величинами.

Вторая предпосылка означает равенство нулю средней величины остатков: $\sum (y - \hat{y}_x)$



в)

Рис 3.4.Зависимость случайных остатков ε_i от теоретических значений y_x

В соответствии с третьей предпосылкой МНК требуется, чтобы дисперсия остатков была гомоскедастичной. Это значит, что для каждого значения фактора x_j остатки ε_i имеют одинаковую дисперсию. Если это условие применения МНК не соблюдается, то имеет место гетероскедастичность. Примеры гетероскедастичности приведены на рис.3.5.

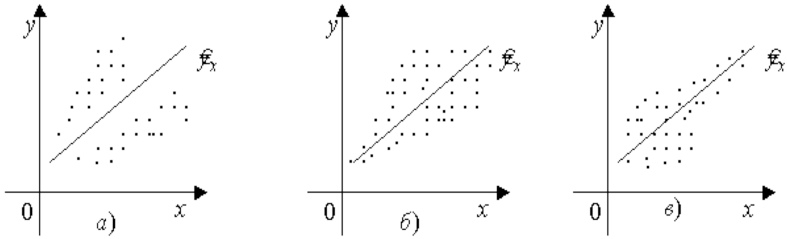


Рис3.5. Примеры гетероскедастичности

Наличие гомоскедастичности или гетероскедастичности можно видеть и по рассмотренному выше графику зависимости остатков от теоретических значений результативного признака y_x.

Для множественной регрессии данный вид графиков является наиболее приемлемым визуальным способом изучения гомо- и гетероскедастичности.

При построении регрессионных моделей чрезвычайно важно соблюдение четвертой предпосылки МНК – отсутствие автокорреляции остатков, т.е. значения ε_i распределены независимо друг от друга.

Автокорреляция остатков означает наличие корреляции между остатками текущих и предыдущих (последующих) наблюдений.

Отсутствие автокорреляции остатков обеспечивает состоятельность и эффективность оценок коэффициентов регрессии.

Предпосылка о нормальном распределении остатков позволяет проводить проверку параметров регрессии и корреляции с помощью критериев t и F. Вместе с тем оценки регрессии, найденные с применением МНК, обладает хорошими свойствами даже при отсутствии нормального распределения остатков.

При несоблюдении основных предпосылок МНК приходится корректировать модель, изменяя ее спецификацию, добавлять (исключать) некоторые факторы и т.д.

При нарушении гомоскедастичности и наличии автокорреляции ошибок рекомендуется традиционный МНК заменять обобщенным МНК.

Обобщенный МНК применяется к преобразованным данным и позволяет получать оценки, которые обладают не только свойством несмещенности, но и имеют меньшие выборочные дисперсии

Тема 4. Информационные технологии в эконометрических исследованиях

Центральной проблемой эконометрики является построение уравнений и систем уравнений (экономико-математических моделей), выражающих экономические закономерности, связи, зависимости, динамические тенденции и определение возможности их практического использования для анализа и прогнозирования.

Построение эконометрических моделей предполагает выполнение множества математических расчетов, обработку больших объемов информации, в связи с чем возникает необходимость в широком использовании компьютерных средств обработки информации. Для этих целей разработаны и широко используются пакеты прикладных программ статистической обработки данных (например, StatGrafics, SPSS, SyStat, Statistica/W. Stadia и др.).

Вследствие большой популярности эконометрических исследований на Западе средства построения эконометрических моделей включены во все известные интегрированные офисные средства (Microsoft Office, Perfect Office и т. д.) и табличные процессоры (Excel, Lotus 1-2-3, Quattro Pro и др.).

Рассмотрим методику построения эконометрических моделей с помощью встроенных функций Microsoft Excel.

В качестве примера рассмотрим следующую задачу. По статистическим данным экономики КР за 1990-2000 гг. требуется построить эконометрические модели, выражающие корреляционную зависимость валового регионального продукта от: численности занятых в экономике и инвестиций. Стоимостные показатели ВРП и инвестиции приводятся в текущих ценах, поэтому они приведены нами к сопоставимому виду на основе темповых показателей и индексов цен производителей.

Построение эконометрических моделей требует выполнения множества расчетов по определению параметров и характеристик.

Все расчеты могут быть выполнены в рамках встроенных статистических функций электронных таблиц.

В зависимости от целей исследования и вида уравнения регрессии расчеты в Excel могут быть выполнены с помощью различных функций ЛИНЕЙН, ЛГРФПРИБЛ, ТЕНДЕНЦИЯ, РОСТ и др.

Таблица 4. 1. 1

Годы	ВРП, млн.сом.	Инвестиции, млн. сом.	Численность занятых, тыс. чел
	Y	X_1	X_2
1990	16409	6569	700,7
1991	14690	10379	691,0
1992	12202	8344	694,4
1993	9969	7852	652,0
1994	8393	5599	653,9
1995	7323	4781	646,7
1996	9924	3490	620,0
1997	6797	2942	655,0
1998	6613	1824	710,0
1999	6795	1470	756,6
2000	7881	1861	754,2

Приведем методику использования MS Excel для построения эконометрических уравнений на примере линейной регрессии (ЛИНЕЙН).

Встроенная статистическая функция ЛИНЕЙН определяет параметры линейной регрессии:

y=mx+b или y=m1x1 + m2x2 +...+ b,

где зависимое значение у является функцией независимого значения *x*. Значения *m* - это коэффициенты, соответствующие каждой независимой переменной *x*, а *b* - константа.

Синтаксис:

ЛИНЕЙН (известные значения *y*; известные значения *x*; конст; статистика)

известные значения *y* - это множество значений *y*, которые уже известны для соотношения *y=mx+b*.

Массив известные значения *x* может содержать одно или несколько множеств переменных.

Конст - это логическое значение, которое указывает, требуются ли, чтобы константа *b* была равна нулю. Константа принимает одно из двух значений ИСТИНА или ЛОЖЬ. Если конст имеет значение истина или опущено, то *b* вычисляется, если конст имеет значение ЛОЖЬ, то *b* полагается равным 0.

Статистика - это логическое значение, которое указывает, требуется ли вернуть дополнительную статистику по регрессии.

Статистика также принимает одно из значений ИСТИНА или ЛОЖЬ. В первом случае дополнительная статистика рассчитывается, во втором случае не рассчитывается.

Дополнительные статистические характеристики функции ЛИНЕЙН приведены ниже Дополнительные статистические характеристики функции ЛИНЕЙН приведены ниже:

$b, m_1, m_2, \ldots m_n$ – коэффициенты регрессии (параметры модели);

$se_1, se_2, \ldots, se_n$ - стандартные значения ошибок для коэффициентов $m_1, m_2, \ldots, m_n$;

se_b - стандартное значение ошибки для постоянной b;

r^2 - коэффициент детерминированности;

se_y - стандартная ошибка для оценки y;

F - F-статистика, используемая для определения того, является ли наблюдаемая взаимосвязь между зависимой и независимой переменными случайной или нет;

df - степени свободы, используемые для нахождения F-критических значений в статистической таблице (для определения уровня надежности модели нужно сравнить значения в таблице с F-статистикой функции ЛИНЕЙН);

ss_{reg} - регрессионая сумма квадратов;

ss_{resid}- остаточная сумма квадратов.

Характеристики выводятся на экран дисплея в виде приведенного ниже массива (таблицы 2):

m_n	m_{n-1}	...	m_2	m_1	B
se_n	Se_{n-1}	...	se_2	se_1	se_b
R^2	Se_y	...			
F	Df	...			
ss_{re} g	ss_{resid}	...			

Порядок выполнения расчетов следующий:

1. Вводятся исходные данные или открывается существующий файл, содержащий исходные данные.

2. В рабочем окне Excel выделяется диапазон ячеек 5*(n+1) (5 число строк, (n+1) - число столбцов, n – число показателей факторов) для вывода результатов расчета.

3. Активизируются "Мастер функций" любым из способов:

а) в главном меню выбирается Вставка/Функция;

б) на панели инструментов Стандартная нажимается кнопка (f_x)

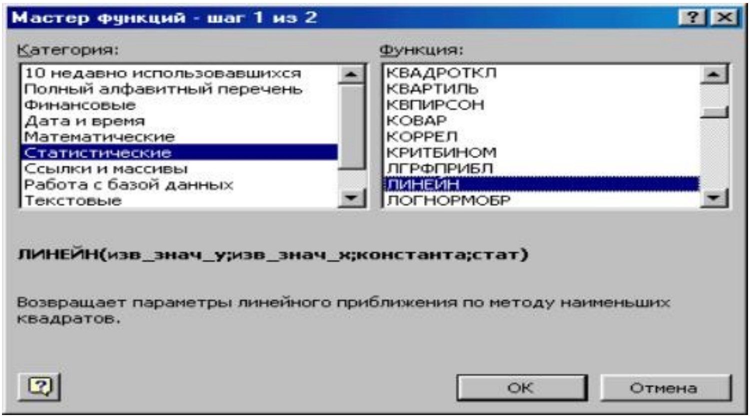


Рис. 4. 1. 1. Диалоговое окно "Мастер функций шаг 1 из 2"

4. В появившемся окне "Мастер функций шаг 1 из 2" среди категорий выбирается Статистические, среди функций - ЛИНЕЙН шаг 1 из 2 (рис. 4.1.1)

5. В появившемся втором окне "Мастер функций" (рис. 4. 1. 2) вводятся аргументы, т.е. указываются диапазоны ячеек рабочего окна EXCEL, в которых находятся исходные данные для *У* и *X*, а также значения аргументов константа и статистика.

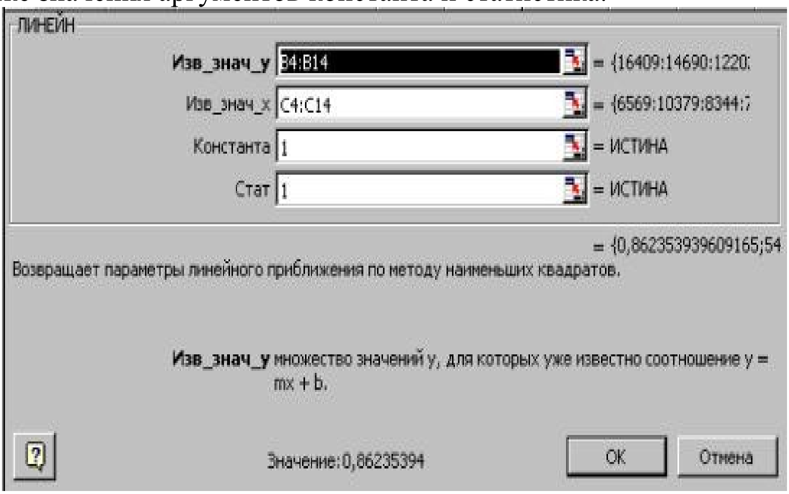


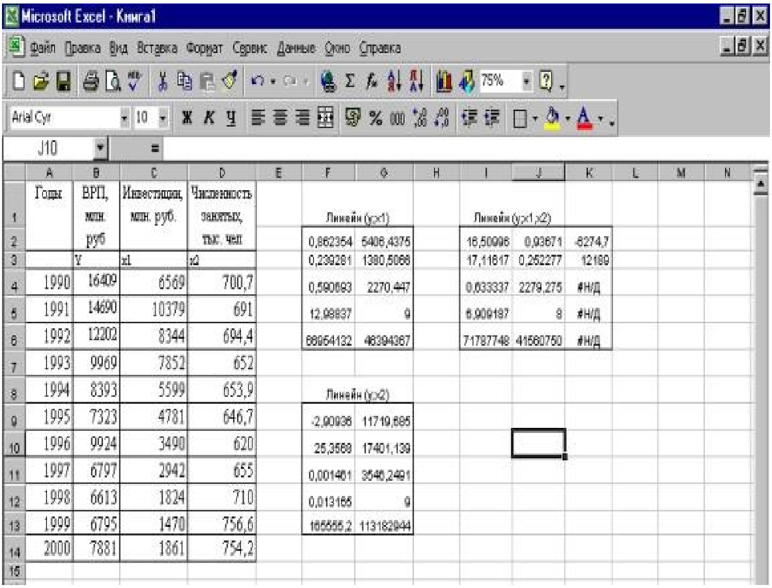
Рис. 4. 1. 2. Второе диалоговое окно "Мастер функций"

6. Нажимается кнопка ОК. В выделенном диапазоне рабочего окна

Excel появляется результат - численное значение для коэффициента регрессии (*b*). Чтобы вывести всю статистику следует нажать клавишу <F2>, а затем - комбинацию клавиш <Ctrl>+<Shift>+<Enter>.

По вышерассмотренным данным (см. Таблица 4.1.1; 4.1.2) получены следующие эконометрические модели:

линейного вида:	экспоненциального вида:
$y=5406{,}43+0{,}86x_1$	$y=6025{,}5 \bullet 1{,}000086^{x1}$
$y= 11719{,}68-2{,}90x_2$	$y=3542{,}4 \bullet 1{,}0014^{x2}$
$y=-6274{,}7+0{,}936x_1+16{,}509x_2$	$y=2147{,}8 \bullet 1{,}000086^{x1} \bullet 1{,}0015^{x2}$



	A	B	C	D	E	F	G	H	I	J	K
1	Годы	ВРП, млн. руб	Инвестиции, млн. руб.	Численность занятых, тыс. чел		Линейн (y;x1)			Линейн (y;x1,x2)		
2						0,862354	5406,4375		16,50996	0,93671	-6274,7
3		Y	x1	x2		0,239281	1380,5066		17,11617	0,252277	12189
4	1990	16409	6569	700,7		0,590693	2270,447		0,633337	2279,275	#Н/Д
5	1991	14690	10379	691		12,98837	9		6,909197	8	#Н/Д
6	1992	12202	8344	694,4		68954132	46394367		71787748	41560750	#Н/Д
7	1993	9969	7852	652							
8	1994	8393	5599	653,9		Линейн (y;x2)					
9	1995	7323	4781	646,7		-2,90936	11719,685				
10	1996	9924	3490	620		25,3569	17401,139				
11	1997	6797	2942	655		0,001461	3546,2491				
12	1998	6613	1824	710		0,013166	9				
13	1999	6795	1470	756,6		165555,2	113182944				
14	2000	7881	1861	754,2							
15											

Рис. 4. 1. 3. Результат вычисления функции ЛИНЕЙН

Тема 5. Системы эконометрических уравнений

5.1. Понятие о системах эконометрических уравнений

Многие экономические взаимосвязи допускают моделирование одним уравнением, но при анализе сложных экономических систем использование отдельных уравнений регрессии является иногда очень грубым предположением: практически изменение одной переменной, как правило, не может происходить при абсолютной неизменности других.

Например, при оценке эффективности производства нельзя руководствоваться только моделью рентабельности. Она должна быть дополнена моделью производительности труда, а также моделью себестоимости единицы продукции.

Поэтому в последние десятилетия в экономических исследованиях важное место заняла проблема описания структуры связей между переменными, системой, так называемых одновременных уравнений, называемых также структурными уравнениями.

Система уравнений в эконометрических исследованиях может быть построена по-разному.

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n \end{cases}$$

Возможна система независимых уравнений, когда каждая зависимая переменная (y) рассматривается как функция одного и того же набора факторов:

Набор факторов x_i в каждом уравнении может варьироваться, например:

$У_1=f(x_1,x_2,x_3,x_4,x_5)$, $y_2=f(x_1,x_3,x_4)$, $y_3=f(x_2,x_4,x_5)$, $y_4=f(x_4,x_5)$ и т.д.

Для определения параметров каждого, отдельно взятого уравнения, используется МНК.

Если зависимая переменная y одного уравнения выступает в качестве фактора x в другом уравнении, то получаем систему рекурсивных уравнений:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nm-1}y_{n-1} + a_{n1}x_1 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

Параметры каждого уравнения данной системы также определяются обычным МНК.

Наибольшее распространение в эконометрических исследованиях получили системы взаимозависимых уравнений, в которых одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других – в правую часть системы:

$$\begin{cases} y_1 = b_{12} \cdot y_2 + b_{13} \cdot y_3 + \ldots + b_{1n} \cdot y_n + a_{11} \cdot x_1 + \ldots + a_{1m} \cdot x_m + \varepsilon_1 \\ y_2 = b_{21} \cdot y_1 + b_{23} \cdot y_3 + \ldots + b_{2n} \cdot y_n + a_{21} \cdot x_1 + \ldots + a_{2m} \cdot x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + \ldots + b_{nn-1} \cdot y_{n-1} + a_{n1} \cdot x_1 + \ldots + a_{nm} \cdot x_m + \varepsilon_n. \end{cases}$$

Такие системы называют также системами совместных, одновременных уравнений. В отличие от предыдущих систем, каждое уравнение данной системы не может рассматриваться самостоятельно. Поэтому для определения параметров этой системы традиционный МНК не приемлем, а используются специальные методы.

Система совместных, одновременных уравнений (которая также называется структурной формой модели) обычно содержит эндогенные и экзогенные переменные. Эндогенные (зависимые) переменные в приведенной ранее системе обозначены как y. Их число равно числу уравнений в системе.

Экзогенные (независимые) переменные, это переменные, влияющие на эндогенные переменные, но не зависящие от них. Обозначаются обычно как x.

Простейшая структурная форма модели имеет вид:

$$\begin{cases} y_1 = b_{12} \cdot y_2 + a_{11} \cdot x_1 + \varepsilon_1; \\ y_2 = b_{21} \cdot y_1 + a_{22} \cdot x_2 + \varepsilon_2. \end{cases}$$

Коэффициенты b^{ij} и a^{ij} - называются структурными коэффициентами модели.

Если пользоваться МНК для оценки структурных коэффициентов модели, то получаются смещенные и несостоятельные оценки. Поэтому для определения структурных коэффициентов модели необходимо: структурную форму модели преобразовать в приведенную форму, которая имеет вид:

$$
\begin{cases}
\hat{y}_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + \ldots\ldots + \delta_{1ж} \cdot x_{ж}; \\
\hat{y}_2 = \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2 + \ldots\ldots + \delta_{2ж} \cdot x_{ж}; \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\hat{y}_{н} = \delta_{н1} \cdot x_1 + \delta_{н2} \cdot x_2 + \ldots\ldots + \delta_{нж} \cdot x_{ж},
\end{cases}
$$

где δ_{ij} - коэффициенты приведенной формы модели.

По своему виду приведенная форма является системой независимых уравнений, поэтому для оценки δ_{ij} можно применить МНК, а затем оценить структурные коэффициенты модели.

5.2. Проблема идентификации модели

В процессе перехода от приведенной формы модели к структурной естественно возникает проблема идентификации, т.е. единственности соответствия между приведенной и структурной формами модели.

Поэтому с позиции идентифицируемости структурные модели можно подразделить на: идентифицируемые, неидентифицируемые и сверхидентифицируемые.

Модель идентифицируема, если все структурные ее коэффициенты определяются однозначно, единственным образом, по коэффициентам приведенной формы, т.е. число параметров структурной модели равно числу параметров приведенной формы модели.

Модель неидентифицируема, если число приведенных коэффициентов меньше числа структурных коэффициентов.

Модель сверхидентифицируема, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы модели можно получить два и более значений одного структурного коэффициента. В такой модели число структурных коэффициентов меньше числа приведенных коэффициентов.

Если обозначить число эндогенных переменных в j-м уравнении системы через H, а число экзогенных переменных, которые содержаться в системе, но не входят в данное уравнение, - через D, то условие идентифицируемости модели запишется следующим образом:

$D+1=H$ – уравнение идентифицируемо,

$D+1<H$ – уравнение неидентифицируемо,

$D+1>H$ – уравнение сверхидентифицируемо.

Рассмотренное правило отражает необходимое, но недостаточное условие идентификации.

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

Идентифицируемая система является сверхидентифицируемой, если содержит хотя бы одно сверхидентифицируемое уравнение.

5.3. Методы оценки параметров одновременных уравнений

Коэффициенты структурной модели могут быть оценены разными методами в зависимости от вида системы одновременных уравнений.

Наиболее из них распространены следующие методы оценки:

косвенный МНК (КМНК);

двухшаговый МНК (ДМНК);

трехшаговый МНК (ТМНК);

метод максимального правдоподобия с полной информацией; метод максимального правдоподобия при ограниченной информации.

Косвенный и двухшаговый МНК рассматриваются как традиционные методы оценки коэффициентов структурной модели.

КМНК применяется для идентифицируемой системы одновременных уравнений, а ДМНК используется для оценки коэффициентов сверхидентифицируемой модели.

Метод максимального правдоподобия рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК, но при большом числе уравнений системы этот метод приводит к достаточно сложным процедурам вычислений.

Дальнейшим развитием ДМНК является ТМНК. Этот метод пригоден для оценки параметров всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективен ДМНК.

Процедура применения косвенного метода наименьших квадратов (КМНК) предполагает выполнение следующих этапов:

преобразование структурной модели в приведенную форму; для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты δ_{ij} приведенной формы модели трансформируются в параметры структурной модели.

Рассмотрим применение КМНК для простейшей идентифицируемой модели:

$$\begin{cases} y_1 = b_{12} \cdot y_2 + a_{11} \cdot x_1 + \varepsilon_1 \\ y_2 = b_{21} \cdot y_1 + a_{22} \cdot x_2 + \varepsilon_2 \end{cases}; \quad (5.1)$$

Приведенная форма модели имеет вид:

$$\begin{cases} y_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + v_1; \\ y_2 = \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2 + v_2; \end{cases} \quad (5.2.)$$

где y_1, y_2 - случайные ошибки приведенной формы.

Применяя, для каждого уравнения этой системы, МНК определяем коэффициенты δ_{ij} .

Соответствующие системы будут выглядеть следующим образом:

$$\begin{cases} \sum y_1 x_1 = \delta_{11} \sum x_1^2 + \delta_{12} \sum x_1 x_2; \\ \sum y_1 x_2 = \delta_{11} \sum x_1 x_2 + \delta_{12} \sum x_2^2; \end{cases} \quad (5.3)$$

$$\begin{cases} \sum y_2 x_1 = \delta_{21} \sum x_1^2 + \delta_{22} \sum x_1 x_2; \\ \sum y_2 x_2 = \delta_{21} \sum x_1 x_2 + \delta_{22} \sum x_2^2; \end{cases} \quad (5.4)$$

Решая системы (5.3), (5.4) находим коэффициенты , , .

При непосредственном применении традиционного МНК к каждому уравнению структурной формы результаты могут сильно отличаться от результатов применения КМНК.

Если система сверхидентифицируема, то КМНК не даст однозначных оценок параметров структурной модели и поэтому он не используется. В этом случае можно использовать разные методы, среди которых наиболее распространен ДМНК.

Основная идея ДМНК – получение на основе приведенной формы модели для сверхидентифицируемого уравнения теоретических значений эндогенных переменных, содержащихся в правой части уравнения.

Затем, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения.

Если все уравнения системы сверхидентифицируемы, 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.

Пусть дана идентифицируемая модель:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}\cdot x_1 + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{22}\cdot x_2 + \varepsilon_2. \end{cases} \qquad (5.5)$$

Если на параметр *b* наложить ограничение, а именно $b_{12} = a_{11}$, то система превращается в простейшую сверхидентифицируемую модель

$$\begin{cases} y_1 = b_{12}(y_2 + x_1) + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{22}\cdot x_2 + \varepsilon_2, \end{cases} \qquad (5.6)$$

в которой 1-е уравнение уже является сверхидентифицируемым:

$H=1$ (y_1), $D=1$ (x_2), значит $D+1>H$.

Второе уравнение является (как и было) точно идентифицируемым:

$H=2$ (y_1, y_2), $D=1$ (x_1), $D+1=H$.

Применим ДМНК к полученной модели (5.6):

<u>1 шаг</u>: применив МНК, найдем приведенную форму модели:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + v_2; \\ y_2 = \delta_{12}x_1 + \delta_{22}x_2 + v_2. \end{cases}$$

<u>2 шаг</u>: на основе 2-го уравнения данной системы, подставляя заданные значения 1_1 и x_2, определяем теоретические значения для эндогенной переменной y_2, т.е. y_2.

Введем новую переменную Z: $Z= y_2+ x_1$.

Тогда 1-е уравнение системы (5.6) $y_1 = b_{12}\cdot Z.$

Применяя МНК к данному уравнению, находим : Σ y Z= ΣZ откуда =Σ y Z/ΣZ .

2-е уравнение системы (5.6) не изменилось, поэтому система одновременных уравнений будет иметь вид:

$$\begin{cases} y_1 = b_{12}(y_2 + x_1); \\ y_2 = b_{21}y_1 + a_{22}\cdot x_2. \end{cases}$$

Приведем некоторые примеры применения эконометрических систем уравнений на практике.

К одной из распространенных систем одновременных уравнений относится статическая модель Кейнса, для описания народного хозяйства страны:

$$\begin{cases} C = a + b \cdot y + \varepsilon, \\ y = C + I, \end{cases}$$

где C – личное потребление, y – национальный доход, I – инвестиции (все в постоянных ценах).

Другим примером может служить динамическая модель Кейнса:

$$\begin{cases} C_t = a + b_1 \cdot Y_t + b_2 \cdot Y_{t-1} + \varepsilon_1, \\ Y_t = C_t + G_t + I_t + L_t, \\ P_t = Y_t + Z_t, \end{cases}$$

где Y^t, C^t, P^t - доход, частное потребление, ВНП, соответственно, в период времени t; Y^{t-1}- доход предыдущего года; G^t, I^t, L^t, Z^t - соответственно, общественное потребление, валовые капиталовложения, изменения складских запасов, сальдо платежного баланса.

Большую популярность получила динамическая модель Клейна.

Система одновременных уравнений нашла применение в исследованиях спроса и предложения. Она имеет вид:

$$\begin{cases} Q^d = a_0 + a_1 P + \varepsilon_1 \\ Q^s = b_0 + b_1 P + \varepsilon_2 \\ Q^d = Q^s \end{cases}$$

где Q, Q - объем спроса и объем предложения, P - цена.

Построение системы структурных уравнений позволяет глубже изучить причины связи результирующих признаков. При этом происходит выделение и оценка косвенных и непосредственных влияний признаков.

Тема 6. Методы и модели анализа динамики экономических процессов

6.1. Понятие экономических рядов динамики. Сглаживание временных рядов

Динамические процессы, происходящие в экономических системах, чаще всего проявляются в виде ряда последовательно расположенных в хронологическом порядке значений того или иного показателя, который в своих изменениях отражает ход развития изучаемого явления в экономике. Предварительно рассмотрим ряд понятий.

Последовательность наблюдений одного показателя (признака), упорядоченных в зависимости от последовательно возрастающих или убывающих значений другого показателя (признака), называют динамическим рядом, или рядом динамики.

Если в качестве признака, в зависимости, от которого происходит упорядочение, выступает время, то такой динамический ряд называется временным рядом.

Эконометрические модели можно строить, используя два типа исходных данных, характеризующих:

совокупность различных объектов в определенный момент (период) времени;

один объект за ряд последовательных моментов (периодов) времени.

Модели, построенные по данным 1-го типа, называются пространственными моделями, по данным второго типа - моделями временных рядов

Время, прошедшее от начального момента наблюдения до конечного, называют длиной временного ряда, а значения показателя в каждый конкретный момент времени - уровнями временного ряда.

Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно разделить на 3 группы:

факторы, формирующие тенденции ряда;

факторы, формирующие циклические колебания ряда;

случайные факторы.

При различных сочетаниях в изучаемом явлении или процессе этих факторов зависимость уровней ряда от времени может принимать различные формы.

Если во временном ряду проявляется длительная тенденция изменения экономического показателя, то говорят, что имеет место тренд (рис.6.1а).

Изучаемый показатель может быть подвержен циклическим колебаниям.

Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей экономики зависит от времени года (например, цены на сельхозпродукты) (рис.6.1б).

Некоторые временные ряды не содержат тенденции и циклической компоненты, а каждый следующий их уровень образуется под воздействием случайной компоненты (например, изменение курса американского доллара по отношению к рублю) (рис. 6.1в).

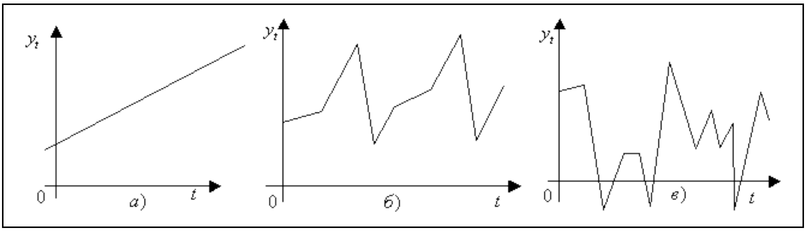


Рис .6.1. Основные компоненты временного ряда:

а)- возрастающая тенденция; б) - сезонная; в) – случайная компонента.

Если модель является временным рядом, представленным как сумма трендовой, циклической и случайной компонент, то такая модель называется аддитивной моделью временного ряда.

Если в модели временный ряд представлен как произведение перечисленных компонент, то такая модель называется мультипликативной моделью временного ряда.

Для статического анализа одномерных временных рядов экономических показателей вида: $y_1, y_2, \dots y_n$ вычисляют ряд величин:

- абсолютный прирост $\Delta y_t = y_t - y_{t-k}$, который показывает величину изменения показателя за определенный интервал времени;

- средний абсолютный прирост: $\Delta \overline{y}_t = \dfrac{y_t - y_{t-k}}{k}$, т.е. прирост в единицу времени;

-коэффициент роста для t-го периода $K_{t(p)} = \dfrac{y_t}{y_t - k}$,

-коэффициент прироста $K_{t(пр)} = K_{t(p)} - 1$.

На практике часто применяют показатель темпа роста и темпа прироста: $T_{t(p)} = \dfrac{y_t}{y_t - k} \cdot 100\%$

где $T_{t(p)}$ - темп роста для t-го периода; $T_{t(пр)} = (T_{t(p)} - 1) \cdot 100\%$,

где $T_{t(пр)}$ - темп прироста для t-го периода.

Важной характеристикой временного ряда является также средний уровень ряда:

$$\bar{y} = \frac{\sum y_t}{n}$$

, где n число наблюдений

Предварительный анализ временных рядов экономических показателей заключается в основном в выявлении и устранении аномальных значений уровней ряда, а также в определении наличия тренда в исходном временном ряде.

Под аномальным уровнем понимается отдельное значение уровня временного ряда, которое не отвечает потенциальным возможностям исследуемой экономической системы и оказывает существенное влияние на значения основных характеристик временного ряда.

Для выявления аномальных уровней временных рядов используются методы, рассчитанные для статистических совокупностей, например, метод Ирвина предполагает использование следующей формулы:

$$\boxed{\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y}};\qquad t = 2,3, \ldots, n.,\qquad \text{где}$$

$$\boxed{\sigma_y = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \bar{y})^2}{n-1}}};\qquad \boxed{\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}}.$$

Расчетные значения λ_2, λ_3 и т.д. сравниваются с табличными значениями критерия Ирвина λ_n, и если какое-то значение оказывается больше табличного, то соответствующее значение y^t уровня ряда считается аномальным.

Для определения наличия тренда в исходном временном ряду применяют ряд методов, в частности метод проверки разностей средних уровней. Реализация этого метода состоит из 4-х этапов:

- на 1-м этапе исходный временной ряд $y_1, y_2, y_3 \dots y_n$ разбивается на две, примерно равные по числу уровней, части: n_1 и n_2 $(n_1 + n_2 = n)$;

на 2-м этапе для каждой из частей вычисляются средние значения и дисперсии:

$$\overline{y_1} = \frac{\sum\limits_{t=1}^{n_1} y_t}{n_1}; \qquad \sigma_1^2 = \frac{\sum\limits_{t=1}^{n_1} (y_t - \overline{y_1})^2}{n_1 - 1};$$

$$\overline{y_2} = \frac{\sum\limits_{t=n_1+1}^{n} y_t}{n_2}; \qquad \sigma_2^2 = \frac{\sum\limits_{t=n_1+1}^{n} (y_t - \overline{y_2})^2}{n_2 - 1}.$$

3-й этап заключается в проверке равенства дисперсий обеих частей с помощью F-критерия Фишера:

$$F = \begin{cases} \dfrac{\sigma_1^2}{\sigma_2^2}, & \text{если} \quad \sigma_1^2 > \sigma_2^2; \\[2ex] \dfrac{\sigma_2^2}{\sigma_1^2}, & \text{если} \quad \sigma_1^2 < \sigma_2^2. \end{cases}$$

Если расчетное значение F меньше $F_{ТАБЛ}$, то переходят к 4-му этапу. Если F $F_{ТАБЛ}$, то делается вывод, что данный метод для определения наличия тренда ответа не дает.

На 4-м этапе определяется расчетное значение критерия Стьюдента по формуле:

$$t = \frac{\left|\overline{y_1} - \overline{y_2}\right|}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

где

$$\sigma = \sqrt{\frac{(n_1 - 1) \cdot \sigma_1^2 + (n_2 - 1) \cdot \sigma_2^2}{n_1 + n_2 - 2}}.$$

Если $t_{рас} < t_{ТАБЛ}$ с заданным уровнем значимости , то тренда нет, в противном случае тренд есть.

Чтобы более четко выявить тенденцию развития исследуемого процесса производят сглаживание (выравнивание) временных рядов.

Сглаживания временных рядов можно осуществлять аналитическими или механическими методами.

Суть аналитических методов заключается в построении кривой, проходящей между конкретными уровнями ряда так, чтобы она отображала тенденцию, присущую ряду, и одновременно освобождала его от незначительных колебаний.

Суть методов механического сглаживания заключается в следующем: берется несколько первых уровней временного ряда, образующих интервал сглаживания, и для них подбирается полином, степень которого должна быть меньше числа уровней, входящих в интервал сглаживания; с помощью полинома определяются новые, выровненные значения уровней в середине интервала сглаживания. Далее интервал сглаживания сдвигается на один уровень ряда вправо, вычисляется следующее сглаженное значение и т.д.

Простейшим методом механического сглаживания является метод простой скользящей средней. Согласно этому методу для временного ряда $y_1, y_2, y_3 \ldots y_n$ определяется интервал сглаживания m ($m<n$). Для первых m уровней временного ряда вычисляется их средняя арифметическая. Затем интервал сглаживания сдвигается на один уровень вправо, повторяется вычисление средней арифметической и т.д.

Для вычисления сглаженных уровней ряда $\overline{y}_t$ применяется формула:

$$\overline{y}_t = \frac{\sum\limits_{t=t-p}^{t+p} y_t}{m}, \quad t > p, \text{ где } \quad p = \frac{m-1}{2}.$$

В результате этой процедуры получаются n-m+1 сглаженных значений уровней ряда; при этом первые p и последние p-уровней ряда теряются.

6.2. Автокорреляционная функция. Коррелограмма

При наличии во временном ряду тенденции и циклических изменений значения последующего уровня ряда зависят от предыдущих. Зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда.

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

Пусть задан временный ряд: $y_1, y_2, y_3 \ldots y_n$ и пусть имеет место линейная корреляция между y_t и y_{t-1}.

Определим коэффициент корреляции между рядами y_t и y_{t-1}.

Для этого воспользуемся следующей формулой:

$$r_{xy} = \frac{\sum(x_j - \overline{x}) \cdot (y_j - \overline{y})}{\sqrt{\sum(x_j - \overline{x})^2 \cdot \sum(y_j - \overline{y})^2}}.$$

Пологая $x_j = y_{t-1}$, $y_j = y_{t-1}$, получим

$$r_1 = \frac{\sum\limits_{t=2}^{n}(y_{t-1} - \overline{y_{t-1}}) \cdot (y_t - \overline{y_t})}{\sqrt{\sum\limits_{t=2}^{n}(y_{t-1} - \overline{y_{t-1}})^2 \cdot \sum\limits_{t=2}^{n}(y_t - \overline{y_t})^2}}, \qquad (6.1)$$

где $\overline{y_t} = \dfrac{\sum\limits_{t=2}^{n} y_t}{n-1}$; $\overline{y_{t-1}} = \dfrac{\sum\limits_{t=2}^{n} y_{t-1}}{n-1}$ (6.2)

Эту величину называют коэффициентом автокорреляции уровней ряда 1-го порядка.

Аналогично определяются коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции 2-го порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле:

$$r_2 = \frac{\sum\limits_{t=3}^{n} (y_{t-2} - \overline{y_{t-2}}) \cdot (y_t - \overline{y_t})}{\sqrt{\sum\limits_{t=3}^{n} (y_{t-2} - \overline{y_{t-2}})^2 \cdot \sum\limits_{t=3}^{n} (y_t - \overline{y_t})^2}}, \qquad (6.3)$$

где $\overline{y_3} = \dfrac{\sum\limits_{t=3}^{n} y_t}{n-2}$; $\overline{y_4} = \dfrac{\sum\limits_{t=3}^{n} y_{t-2}}{n-2}$ (6.4)

Порядок уровня ряда автокорреляции называют лагом.

Для формулы (6.1) лаг равен единице, для (6.3) –двум.

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

График зависимости ее значений от величины лага называется коррелограмой.

АКФ и коррелограмма позволяют определить лаг, при котором автокорреляция наиболее высокая, а, следовательно, и лаг, при котором связь между текущим и предыдущим уровнями ряда наиболее тесная, т.е. с их помощью можно выявить структуру ряда.

Коэффициент автокорреляции и АКФ целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты и циклической компоненты:

если наиболее высоким оказался коэффициент автокорреляции 1-го порядка, то исследуемый ряд содержит только тенденцию;

если наиболее высоким оказался коэффициент автокорреляции к-го порядка, то ряд содержит циклические колебания с периодичностью в к-моментов времени;

если, ни один из коэффициентов не является значимым, то можно сделать одно из двух предположений, относительно структуры этого ряда: либо ряд не содержит тенденции и циклических изменений и имеет структуру, сходную со структурой ряда, изображенного на рис.6.1в, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

6.3. Автокорреляция в остатках. Критерий Дарбина-Уотсона

Рассмотрим уравнение регрессии вида:

$$\hat{y}_t = a + \sum_{j=1}^{k} b_j \cdot x_{jt} + \varepsilon_t$$

где k – число независимых переменных модели.

Для каждого момента (периода) времени t=1: n значение компоненты ε_t определяется так: $\varepsilon_t = y_t - \hat{y}_t$,

или $$\boxed{\varepsilon_t = y_t - (a + \sum_{j=1}^{k} b_j \cdot x_{jt})}$$.

При моделировании временных рядов нередко встречается ситуация, когда остатки ε_t содержат тенденцию (рис.6.2.б,в) или циклические колебания (рис.6.2.г), когда в соответствии с предпосылками МНК остатки должны быть случайными (рис. 6.2 а).

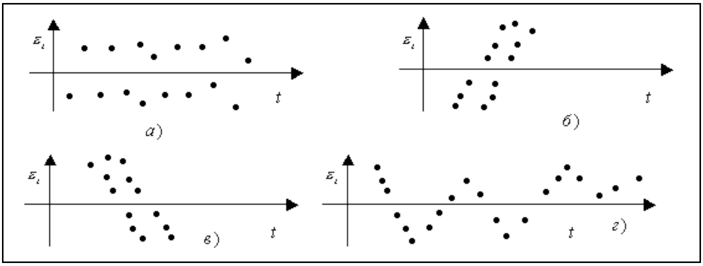


Рис.6.2 Варианты изменения остатков временного ряда

В том случае, когда каждое следующее значение ε_t зависит от ε_{t-1}, говорят о наличии автокорреляции остатков.

Причинами автокорреляции могут быть:

- исходные данные с ошибками в измерениях результативного признака;
- формулировка модели (модель может не включать фактор, оказывающий существенное воздействие на результат). Очень часто этим фактором является фактор времени *t*).

Если причина автокорреляции – в неправильной спецификации функциональной формы модели, то следует изменить форму связи факторных и результативных признаков.

Существуют два наиболее распространенных метода определения автокорреляции остатков:

1) путем построения графика зависимости остатков ε_t от времени и визуальное определение наличия или отсутствия автокорреляции;

2) использование критерия Дарбина-Уотсона и расчет величины

$$\boxed{d = \frac{\sum\limits_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum\limits_{t=1}^{n} \varepsilon_t^2}}.$$

Из данной формулы нетрудно вывести следующее соотношение между критерием Дарбина-Уотсона и коэффициентом автокорреляции остатков первого порядка

$\boxed{d \approx 2 \cdot (1 - r_1^{\varepsilon})}$, где $\boxed{r_1^{\varepsilon} = \dfrac{\sum(\varepsilon_t - \bar{\varepsilon}_t)(\varepsilon_{t-1} - \bar{\varepsilon}_{t-1})}{\sqrt{\sum(\varepsilon_t - \bar{\varepsilon}_t)^2 \sum(\varepsilon_{t-1} - \bar{\varepsilon}_{t-1})^2}}}$

Иными словами, если в остатках существует полная положительная автокорреляция и $r_1^{\varepsilon} = 1$, то d=0; если в остатках полная отрицательная автокорреляция $r_1^{\varepsilon} = -1$, то, d=4; если автокорреляция остатков отсутствует, то $r_1^{\varepsilon} = 0$ и d=2.

Следовательно, $0 \le d \le 4$

Есть несколько существенных ограничений на применение критерия Дарбина – Уотсона:

- он непременим к модели авторегрессии;
- данный критерий можно использовать только для выявления автокорреляции остатков 1-го порядка;
- критерий дает достоверные результаты только для больших выборок.

6.4. Моделирование тенденций временного ряда. Адаптивные модели прогнозирования

Одним из наиболее распространенных способов моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени или тренда. Этот способ называют аналитическим выравниванием временного ряда.

Для построения трендов чаще всего применяются следующие функции:

линейный тренд: $\hat{y} = a + b \cdot t$;

гипербола: $\hat{y} = a + \dfrac{b}{t}$;

экспоненциальный тренд: $\hat{y} = e^{a+b\cdot t}$ *или* $\hat{y} = a \cdot b^t$

полиномиальный тренд: $\hat{y} = a + b_1 \cdot t + b_2 \cdot t^2$ - полином 2-й степени;

$\hat{y} = a + b_1 \cdot t + b_2 \cdot t^2 + b_3 \cdot t^3$ - полином 3-й степени.

Параметры каждого из перечисленных трендов можно определить обычным МНК.

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 целях можно использовать коэффициенты автокорреляции уровней ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни y^t и y_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции 1-го порядка уровней исходного ряда должен быть высоким.

Выбор наилучшего уравнения в случае, если ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации $\bar{R}$ и выбора уравнения тренда с максимальным значением $\bar{R}$. Реализация этого метода относительно проста при компьютерной обработке данных.

При наличии неявной нелинейной тенденции следует дополнять описанные выше методы качественным анализом динамики изучаемого показателя, с тем, чтобы избежать ошибок спецификации при выборе вида тренда.

Качественный анализ предполагает изучение проблем возможного наличия в исследуемом временном ряде поворотных точек и изменения темпов прироста, начиная с определенного момента. В случае если уравнение тренда выбрано неверно при больших значениях *t*, результаты прогноза на основе выбранного вида тренда будут недостоверными.

Существует несколько подходов к анализу структуры временных рядов, содержащих сезонные или циклические колебания.

Рассмотрим простейший подход – построение аддитивной или мультипликативной модели временного ряда методом скользящей средней.

Аддитивная и мультипликативная модели математически записываются соответственно

$$\boxed{Y = T + S + E,} \qquad (6.5)$$

$$Y = T \cdot S \cdot E \qquad (6.6)$$

где *T, S, E* – трендовая, сезонная и случайная компоненты, соответственно.

Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, если она возрастает или уменьшается, строят мультипликативную модель временного ряда.

Процесс построения модели включает следующие шаги:

выравнивание исходного ряда методом скользящей средней;

расчет значений сезонной компоненты S;

устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных ($\boxed{T + E}$) в аддитивной или ($\boxed{T \cdot E}$) в мультипликативной модели;

аналитическое выравнивание уровней ($\boxed{T + E}$) или ($\boxed{T \cdot E}$) и расчет значений T с использованием полученного уравнения тренда;

расчет полученных по модели значений ($\boxed{T + E}$) или ($\boxed{T \cdot E}$);

расчет абсолютных и относительных ошибок.

Другой метод моделирования временного ряда, содержащего сезонные изменения, состоит в построении модели регрессии с включением фактора времени и фиктивных переменных. Количество фиктивных переменных должно быть на единицу меньше числа периодов внутри одного цикла колебаний.

В основе экстраполяционных методов прогнозирования лежит предположение о том, что основные факторы и тенденции, имевшие место в прошлом, сохраняются в будущем.

При краткосрочном прогнозировании, а также при прогнозировании в ситуации изменения внешних условий, когда более важными являются последние реализации исследуемого процесса, более эффективными оказываются адаптивные методы, учитывающие неравноценность уровней временного ряда.

Адаптивные модели прогнозирования – это модели дисконтирования данных, способные быстро приспосабливать свою структуру и параметры к изменению условий. Инструментом прогноза в адаптивной модели является математическая модель, аргументом которой выступает – время.

При оценке параметров адаптивных моделей, в отличии от «кривых роста», наблюдениям (уровням ряда) присваиваются различные веса, в зависимости от того, насколько сильным признается их влияние на текущий уровень. Это позволяет учитывать изменения в тенденции, а также любые колебания, в которых прослеживается закономерность.

Тема 7. Моделирование динамических процессов

7.1. Характеристика моделей с распределенным лагом и моделей авторегрессии

Многие экономические процессы имеют временной характер изменения. Поэтому в эконометрике рассматриваются так называемые динамические модели.

Эконометрическая модель является динамической, если в данный момент времени t она учитывает значения входящих в нее переменных, относящихся как к текущему, так и к предыдущим моментам времени.

Выделяют два основных типа динамических эконометрических моделей.

К **моделям 1-го типа** относятся модели авторегрессии и модели с распределенным лагом, в которых значения переменной за прошлые периоды времени (лаговые переменные) непосредственно включены в модель. **Модели 2-го типа** учитывают динамическую информацию в неявном виде. В этих моделях присутствуют переменные, характеризующие ожидаемый или желаемый уровень результата, или одного из факторов в момент времени t. Этот уровень считается неизвестным, и определяются экономическими единицами с учетом информации, которой они располагают в момент (t-1).

В зависимости от способа определения ожидаемых значений показателей различают модели неполной корректировки, адаптивных ожиданий и рациональных ожиданий. Оценка параметров этих моделей сводится к оценке параметров моделей авторегрессии.

При исследовании экономических процессов нередко приходится моделировать ситуации, когда значение результативного признака в момент времени t формируется под воздействием ряда факторов, действовавших в прошлые моменты времени (t-1),(t-2),..(t-k). Например, выручка от реализации и прибыль компании текущего периода зависят от расходов на рекламу или проведенных маркетинговых исследований, сделанных компаний в предшествующие моменты времени.

Величину k, характеризующую запаздывание в воздействии фактора на результат, называют в эконометрике **лагом**, а временные ряды самых факторных переменных, сдвинутые на один или более моментов времени, - **лаговыми переменными.**

Основная проблема экономической политики как на макро-, так и на микроуровне это решение обратного типа задач, т.е. задач, определяющих, какое воздействие окажут значения управляемых переменных текущего периода на будущие значения экономических показателей. Например, как повлияют инвестиции в промышленность на валовую добавленную стоимость этой отрасли экономики будущих периодов.

Модели экономических процессов, содержащие не только текущие, но и лаговые значения факторных переменных называются **моделями с распределенным лагом.**

Например, модель

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + \varepsilon_t$$

является моделью с распределенным лагом.

Если на величину зависимой переменной текущего периода (y_t) оказывают влияние ее значения в прошлые моменты времени (y_{t-1}, y_{t-2},…), то эти процессы обычно описываются с помощью **моделей авторегрессии.** Например,

$$y_t = a + b_0 x_t + c_1 y_{t-1} + c_2 y_{t-2} + \varepsilon_t.$$

Построение моделей с распределенным лагом и моделей авторегрессии имеет свою специфику:

оценка параметров моделей авторегрессии, а в большинстве случаев и моделей с распределенным лагом не может быть произведена с помощью обычного МНК и требует специальных статистических методов;

приходится решать проблему выбора оптимальной величины лага и определения его структуры;

между моделями с распределенным лагом и моделями авторегрессии существует определенная взаимосвязь, и в некоторых случаях необходимо осуществлять переход от одного типа моделей к другому.

Рассмотрим модель с распределенным лагом в ее общем виде:

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + \ldots + b_p \cdot x_{t-p} + \varepsilon_t. \qquad (7.1)$$

Коэффициент b_0 характеризует среднее абсолютное изменение y_t при изменении x_t на единицу в момент времени t, без учета воздействия лаговых значений фактора x. Этот коэффициент называют **краткосрочным мультипликатором.**

В момент (t+1) совокупное воздействие факторной переменной x^t на результат y^t составит (b_0+b_1) условных единиц, в момент (t+2) это воздействие можно охарактеризовать суммой (b_0+b_1+b_2) и т. д..

Полученные таким образом новые коэффициенты называют **промежуточными мультипликаторами.**

Введем обозначение

$$b_0 + b_1 + \ldots + b_\ell = b$$

Величину b называют **долгосрочным мультипликатором**. Он показывает абсолютное изменение в долгосрочном периоде $t + \ell$ результата y под влиянием изменения на единицу фактора x.

Положим

$$\beta_j = \frac{b_j}{b}, \quad j = \overline{0,\ell}$$

Полученные величины называются **относительными коэффициентами** модели с распределенным лагом.

Очевидно, что если $0 < \beta_j < 1$ и $\sum_{j=0}^{\ell} \beta_j = 1$

В этом случае относительные коэффициенты β_j являются весами для соответствующих коэффициентов b_j. Каждый из них изменяет долю общего изменения результативного признака в момент времени ($t+j$). На основе величин β_j можно определить две важные характеристики модели множественной регрессии: величину среднего и медианного лага.

Средний лаг определяется по формуле $\overline{\ell} = \sum_{j=0}^{\ell} j \cdot \beta_j$ и представляет собой средний период, в течение которого будет происходить изменение результата под воздействием изменения фактора в момент времени t. Небольшая величина среднего лага свидетельствует об относительно быстром реагировании результата на изменение фактора, высокое его значение о том, что воздействие фактора на результат будет сказываться в течение длительного периода времени.

Медианный лаг - это величина лага, для которого $\sum_{j=0}^{l_{me}} \beta_j \approx 0{,}5$.

Это тот период времени, в течение которого будет реализована половина общего воздействия фактора на результат.

Рассмотрим теперь модель авторегрессии:

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \varepsilon_t \qquad (7.2)$$

Параметров b^0 имеет тот же смысл, что и в модели с распределенным лагом (1).Общее абсолютное изменение результата в момент (t+1) составит $b_0 \cdot c_1$, в момент времени (t+2) абсолютное изменение результата составит $b_0 \cdot c_1^2$ единиц и т.д.

Долгосрочный мультипликатор в модели авторегрессии рассчитывается как сумма как сумма краткосрочного и промежуточных мультипликаторов:

$$b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 + \dots \qquad (7.3)$$

Обычно во всех моделях авторегрессии вводится условие стабильности, состоящее в том, что коэффициент регрессии при переменной y_{t-1} по абсолютной величине меньше единицы ($|c_1|<1$).

Следовательно, можно написать:

$$b = b_0 \cdot (1 + c_1 + c_1^2 + c_1^3 + \dots) = \frac{b_0}{1 - c_1}. \qquad (7.4)$$

7.2. Модели адаптивных ожиданий и неполной корректировки

Эконометрические методы активно используются также и в макроэкономических моделях, в которых учитываются ожидания экономических агентов относительно значений экономических показателей, включенных в модель, в момент времени t.

В зависимости от гипотезы о механизме формирования этих ожиданий различают модели адаптивных ожиданий, неполной корректировки и рациональных ожиданий.

Модели рациональных ожиданий достаточно сложны и требуют специальных методов математической статистики. Поэтому ограничимся рассмотрением адаптивных ожиданий и неполной корректировки.

Модели адаптивных ожиданий.

Рассмотрим модель вида

$$y_t = a + b \cdot x_{t+1}^* + \varepsilon_t, \qquad (7.5)$$

где y^t - фактическое значение результативного признака;

x^*_{t+1} - ожидаемое значение факторного признака.

Механизм формирования ожиданий в этой модели следующий:

$$x^*_{t+1} - x^*_t = \alpha \cdot (x_t - x^*_t) \quad \text{или} \quad x^*_{t+1} = \alpha \cdot x_t + (1-\alpha) \cdot x^*_t, \qquad (7.6)$$

где $0<\alpha<1$, x^t - фактическое значение.

Из (7.6) следует, что в каждый период времени t+1 ожидания корректируются на некоторую долю α. Параметр α называется **коэффициентом ожиданий.**

Чем ближе α к 1, тем в большей степени реализуются ожидания экономических агентов. И, если α приближается к 0, то это свидетельствует об устойчивости существующих тенденций.

Если выражение (7.6) подставить в (7.5), то получим

$$y_t = a + \alpha \cdot b \cdot x_t + (1-\alpha) \cdot b \cdot x^*_t + \varepsilon_t \qquad (7.7)$$

На основе (7.15) для периода (t-1) можно получить:

$$y_{t-1} = a + \alpha \cdot b \cdot x_{t-1} + (1-\alpha) \cdot b \cdot x^*_{t-1} + \varepsilon_{t-1} \qquad (7.8)$$

Умножив (7.8) на (1-α) и вычитывая из (7.7) можно получить

$$y_t = \alpha \cdot a + \alpha \cdot b \cdot x_t + (1-\alpha) \cdot b \cdot y_{t-1} + U_t, \qquad (7.9)$$

где $U_t = \varepsilon_t - (1-\alpha) \cdot \varepsilon_{t-1}$

Т.о. получили модель авторегрессии, определив параметры которой можно легко перейти к исходной модели (7.9).

Модель (7.9) включает только фактические значения переменных (x_t, y_t, y_{t-1}).

Модель (7.6) называется долгосрочной функцией модели адаптивных ожиданий. Модель (7.9) – краткосрочной функцией модели адаптивных ожиданий.

Модель неполной корректировки. В модели неполной корректировки предполагается, что уравнение определяет не фактическое значение зависимой переменной y_t, а ее ожидаемый (желаемый) уровень y^*_t

$$y_t^* = a + b \cdot x_t + \varepsilon_t \qquad (7.10)$$

Предполагается, что фактическое приращение зависимой переменной $(y^t - y^{t-1})$ пропорционально разнице между ее желаемым уровнем и значением в предыдущий период,

$$(y^t - y^{t-1}) = \lambda \cdot (y_t^* - y_{t-1}) + v_t$$

Откуда $y_t = \lambda \cdot y_t^* + (1 - \lambda) \cdot y_{t-1} + v_t \qquad (7.11)$

Т.о., фактическое значение результата текущего периода y_t есть средняя арифметическая взвешенная его ожидаемого значения текущего периода y_t^* и фактического значения за предыдущий период времени y^{t-1}. Чем больше значение λ, тем быстрее происходит процесс корректировки. Если $\lambda = 1$, то $y_t = y_t^*$ и полная корректировка происходит за один период. Если $\lambda = 0$, то корректировка не происходит вообще.

Подставляя уравнение (7.10) в найденное выражение (7.11), можно получить:

$$y_t = \lambda \cdot a + \lambda \cdot b \cdot x_t + (1 - \lambda) \cdot y_{t-1} + U_t, \qquad (7.12)$$

где $U_t = \beta \cdot \varepsilon_t + v_t$

Соотношение (7.12) есть основное уравнение модели неполной корректировки.

Легко заметить, что уравнение (7.12) включает только фактические значения переменных. Зная оценки параметров этого уравнения, можно определить λ. Затем путем алгебраических преобразований рассчитываются параметры a и b уравнения (7.10). Уравнение (7.10) называют также долгосрочной функцией модели неполной корректировки.

Использованные источники

1. Доугерти К. Введение в эконометрику:- Учебник для вузов, М.: Инфа-М, 2001.-402с.

2. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов. – М.: ЮНИТИ- ДАНА, 2002. - 311с.

3. Эконометрика. Учебник для вузов. Под ред. И.И. Елисеевой: – М.: Финансы и статистика, 2001.

4. Бородич С.А. Эконометрика: Учебное пособие. – Мн.: Новое знание, 2001.- 408с.

5. Колемаев В.А. Математическая экономика. - М.: ЮНИТИ,1998.

Содержание

Акжолова Мейманбу Жолдошовна

Курс лекций по

ЭКОНОМЕТРИКЕ

Редактор: Иманова А.

Компьютерная верстка

и макетирование: Салаев Ж.

Подписано к печати 05.11.07
Формат 60х84 1.16 Офсетная печать.
Объем 8,3 п.л. Тираж 100 экз.

Отпечатано в издательском центре ЖАГУ
715600 г.Жалалабат, ул.Ленина, 57