МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

Институт экономики и финансов

Кафедра статистики, эконометрики и естествознания

Е. И. КАДОЧНИКОВА

ЭКОНОМЕТРИКА

Конспект лекций

Казань – 2014

Направление подготовки: 080100.62 «Экономика», бакалавриат, заочное

Учебный план:

«Бухгалтерский учет, анализ и аудит» второе высшее 2013г.,

«Финансы и кредит» второе высшее 2013г.,

«Экономика предприятий и организаций» второе высшее, 2013г.

Дисциплина: Б. 3. «Эконометрика», 1 курс, форма контроля: зачет

Количество часов: 212 (в том числе: лекции - 8, практические занятия - 12, самостоятельная работа – 192)

Аннотация: Целью дисциплины «Эконометрика» является обучение студентов теоретическим основам эконометрической методологии и практическим навыкам применения эконометрических методов для исследования экономических закономерностей и взаимосвязей между экономическими переменными. Данный курс включен в раздел Б3.Б3. ОПД.Ф.3 профессионального цикла дисциплин и относится к базовой части. Изучению дисциплины «Эконометрика» предшествует освоение следующих дисциплин: «Математический анализ», «Теория вероятностей и математическая статистика», «Линейная алгебра», «Микроэкономика», «Макроэкономика», «Статистика». В круг основных задач дисциплины «Эконометрика» входят: получение теоретических знаний об эконометрических методах эмпирического анализа экономических процессов с целью имитации альтернативных сценариев развития анализируемой системы; формирование умения выбирать необходимые инструменты эконометрического анализа для обоснования управленческих решений; обучение практическим навыкам использования эконометрических подходов к моделированию социально–экономических процессов и содержательного обоснования полученных результатов для принятия управленческих решений. Подготовленный материал можно изучать самостоятельно, выполняя предлагаемые задания и проводя самоконтроль усвоения материала с помощью вопросов для самоконтроля.

Темы: 1. Эконометрика как научная дисциплина. 3. Линейная модель парной регрессии и метод наименьших квадратов (МНК). 4. Экономическая и статистическая интерпретация линейной модели парной регрессии. 5,6. Линейная модель множественной регрессии, оценка ее параметров. 8,9. Гетероскедастичность и автокорреляция в остатках регрессии. 15. Модели одномерных временных рядов. 17. Стационарные и нестационарные временные ряды. 19. Понятие о системах эконометрических уравнений. 20. Методы оценки систем одновременных уравнений.

Ключевые слова: *метод наименьших квадратов, линейная модель регрессии, гетероскедастичность, автокорреляция, динамические модели, система взаимозависимых уравнений, косвенный метод наименьших квадратов.*

Дата начала использования: 1 ноября 2012 г.

Автор: Кадочникова Екатерина Ивановна, кандидат экономических наук, доцент кафедры статистики, эконометрики и естествознания ЭиФ КФУ, e-mail: kad-ekaterina@yandex.ru

URL электронного курса в MOODLE:
http://tulpar.kpfu.ru/course/view.php?id=382

Оглавление

Тема 1. Эконометрика как научная дисциплина

Лекция 1(1)

Вопросы для изучения:

1. Цели, предмет, задачи эконометрики.
2. Инструментарий эконометрики. Типы моделей и переменных.
3. Этапы эконометрического моделирования

Аннотация. Данная тема раскрывает основные понятия эконометрики.

Ключевые слова. Модели, переменные, типы данных, этапы моделирования.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где даются общие представления по теме.
- В качестве самостоятельной работы предлагается ознакомиться с решениями типовых задач, презентацией и ответить на вопросы для изучения.
- Для проверки усвоения темы имеется тест для самоконтроля.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=382

2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none)С. 4-7.

3. Эконометрика: учебник / И. И. Елисеева. – М.: Проспект, 2010. – 288 с. С. 6-11.

Цели, предмет, задачи эконометрики

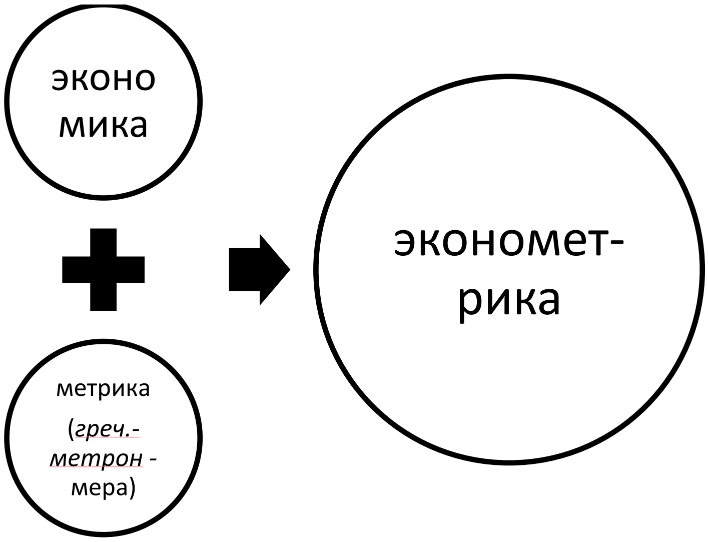


Рис. 1.1. Термин «эконометрика»

Эконометрика – это наука, в которой на базе реальных статистических данных строятся, анализируются и совершенствуются математические модели реальных экономических явлений. Эконометрика позволяет найти количественное подтверждение либо опровержение того или иного экономического закона либо гипотезы.

Термин «эконометрика» впервые был использован бухгалтером П. Цьемпой в Австро-Венгрии, в 1910 году.

1912 г. – И. Фишер сделал безуспешную попытку создать группу ученых для стимулирования развития экономической теории путем ее связи со статистикой и математикой.

1930 г., 29 декабря – на заседании Американской ассоциации развития науки по инициативе И. Фишера, Й. Шумпетера, О. Андерсона , Я. Тинбергена создано эконометрическое общество, на котором норвежский ученый Р. Фриш дал новой науке название «эконометрика».

1933 г. – стал издаваться журнал «Econometrica».

1941 г. – издан первый учебник по эконометрике, автор Я. Тинберген.

1970 – е гг. – противоречия между кейнсианцами, монетаристами и марксистами привели к тому, что методы эконометрики стали применяться не только для оценки теоретических моделей, но и для доказательства причинности при выборе теоретических концепций. Появление компьютеров, создание ARIMA-моделей, VAR-моделей, развитие анализа временных рядов.

Определения эконометрики:

«Эконометрика – это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов, которые раскрыты и обоснованы экономической теорией» (И.И. Елисеева, см. 1, стр. 16).

«Эконометрика – это наука, которая на базе экономической теории, экономической статистики, экономических измерений и математико-статистического инструментария придает количественное выражение качественным закономерностям, обусловленным экономической теорией» (С. А. Айвазян, см. 3, стр. 12).

«Эконометрика – это наука, связанная с эмпирическим выводом экономических законов» (Магнус Я. Р., см. 6., стр. 13).

«Эконометрика – это наука, в которой на базе реальных статистических данных строятся, анализируются и совершенствуются математические модели реальных экономических явлений» (С. А. Бородич, см. 2, стр. 7).

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эконометрика как научная дисциплина зародилась и получила развитие на основе слияния экономической теории, математической экономики и экономической и математической статистики.

«Эконометрика – это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией. Эконометрика не является синонимом приложений математики к экономике. Каждая из трех отправных точек – статистика, экономическая теория и математика – необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. Это единство всех трех составляющих. И это единство образует эконометрику» (Р. Фриш, 1933 г., см. 1, стр. 16).

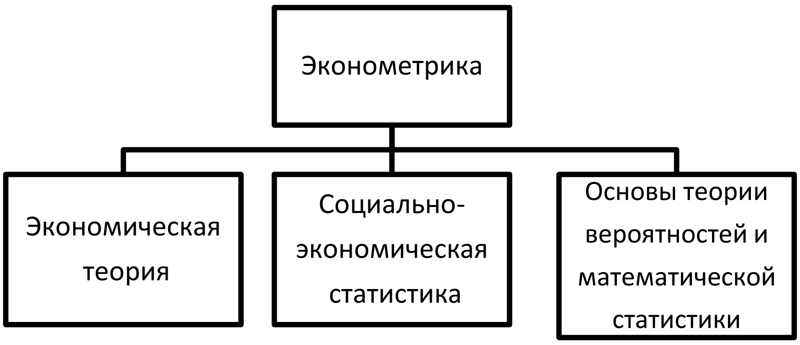


Рис. 1.2. Три составляющие эконометрики

Предметом эконометрики являются количественные закономерности между экономическими явлениями. Однако, в отличие от экономической теории, эконометрика делает упор на количественные, а не на качественные аспекты этих явлений. Например, известно, что спрос на товар с ростом его цены падает. Однако, как быстро и по какому закону это происходит, в экономической теории не определяется. Это в каждом конкретном случае делает эконометрика. С другой стороны, математическая экономика строит и анализирует модели экономических процессов без использования реальных числовых значений. Эконометрика же изучает модели на базе эмпирических данных. Наконец, в эконометрике широко используется аппарат математической статистики, особенно при установлении связей между экономическими показателями. В то же время в экономике невозможно проведение управляемого эксперимента, и эконометристы используют свои собственные приемы анализа, которые в математической статистике не встречаются.

Основными целями эконометрики являются:

1. Прогноз экономических и социально-экономических показателей, характеризующих состояние и развитие анализируемой системы.

2. Имитация различных возможных сценариев социально-экономического развития.

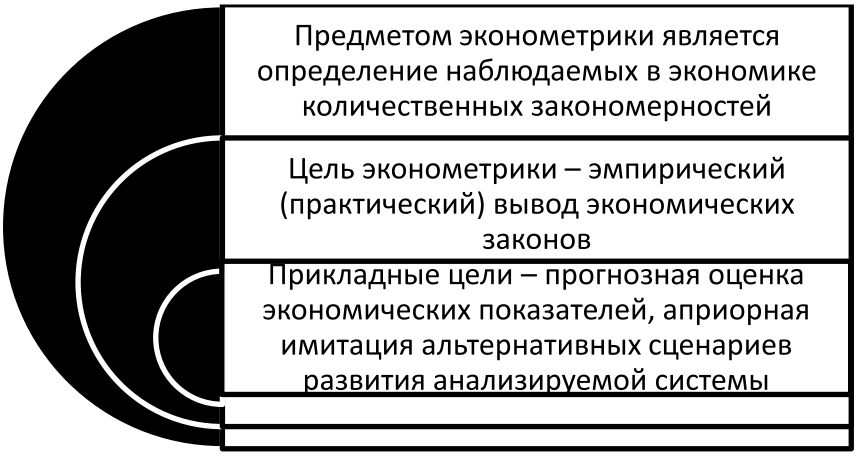


Рис. 1.3. Предмет и цели эконометрики

Основные задачи эконометрики:

- построение эконометрической модели;

- оценка параметров построенной модели, делающих выбранную модель наиболее адекватной реальным данным;

- проверка качества найденных параметров модели и самой модели в целом;

- использование построенных моделей для объяснения поведения исследуемых экономических показателей, прогнозирования, осмысленного проведения экономической политики (С. А. Бородич).

Инструментарий эконометрики. Типы моделей и переменных. Инструментарий эконометрики включает четыре основных раздела: линейная модель регрессии и МНК; обобщенная линейная модель регрессии и ОМНК; статистический анализ временных рядов; анализ систем одновременных уравнений.

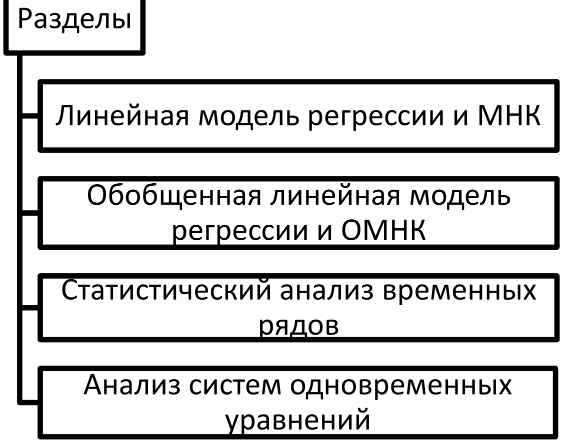


Рис. 1.4. Разделы эконометрики

Особенности эконометрического метода заключаются в следующем:

- исследование статистических зависимостей, а не функциональных;
- отражение особенностей экономических переменных и связей между ними (оптимальность и взаимодействие переменных);
- содержательное обоснование уравнений;
- изучение всей совокупности связей между переменными, а не изолированно взятого уравнения регрессии;
- развитие анализа временных рядов через решение проблем ложной корреляции, лага и других.

Для моделирования эконометрических взаимосвязей между экономическими явлениями чаще всего применяется три типа моделей и три типа переменных.

Рис. 1.5. Типы моделей

Рис. 1.6. Типы переменных

Рис. 1.7. Различия переменных

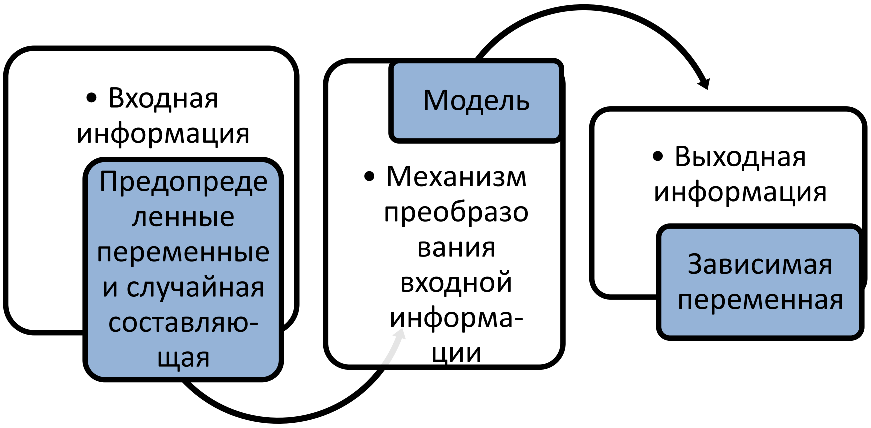


Рис. 1.8. Взаимодействие переменных

Приведем примеры эконометрических моделей.

Модели временных рядов:

- модель тренда: $Y_t = T_t + \varepsilon_t$

- модель сезонности: $Y_t = S_t + \varepsilon_t$

- модель тренда и сезонности: $Y_t = T_t + S_t + \varepsilon_t;$
$Y_t = T_t \cdot S_t \cdot \varepsilon_t$

Модель регрессии:

$$Y_x = f(x, \beta) = f(x_1, \ldots x_k, \beta_1 \ldots \beta_k)$$

Линейная модель множественной регрессии:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon$$

Система одновременных уравнений:

$$\begin{cases} y_1 = b_{12} y_2 + a_{11} x_1 + a_{12} x_2 + a_{13} x_3 + \varepsilon_1, \\ y_2 = b_{21} y_1 + a_{21} x_1 + a_{22} x_2 + a_{23} x_3 + \varepsilon_2 \end{cases}$$

В эконометрике применяют три типа исходных данных.

Множество данных, состоящих из наблюдений за несколькими однотипными статистическими объектами в течение одного периода или за один момент времени, называется перекрестными данными.

Множество данных, состоящих из наблюдений за одним статистическим объектом в течение нескольких периодов или за несколько моментов времени, называется временным рядом.

Множество данных, состоящих из наблюдений за несколькими однотипными статистическими объектами в течение нескольких временных периодов, называется панельными, или пространственными, данными.

Этапы эконометрического моделирования. Этапы эконометрического моделирования:

1. постановочный – определение целей и задач модели;
2. априорный – предварительный анализ ситуации;
3. спецификация модели – выбор типа модели, состава переменных и формы математической связи между ними;
4. информационный – сбор первичной информации;
5. идентификация модели – оценивание параметров модели;
6. верификация модели– проверка качества модели в целом и ее параметров;
7. интерпретация результатов – формулирование выводов и рекомендаций на основе построенной модели.

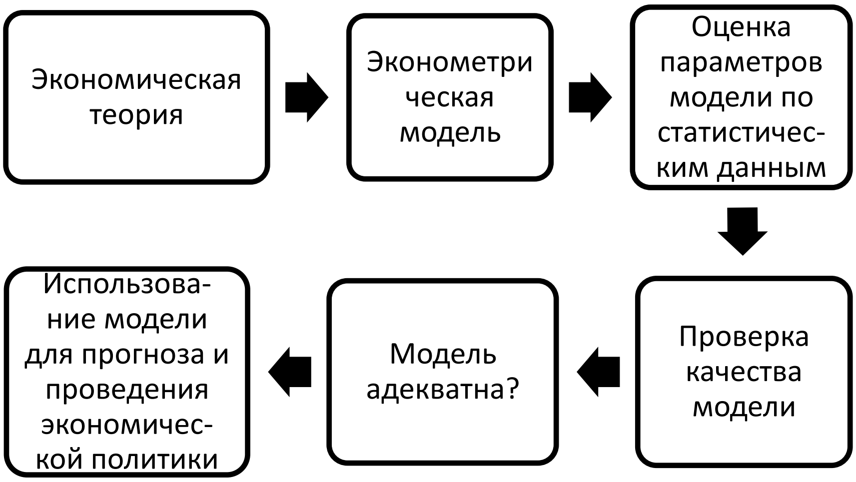


Рис. 1.9. Этапы эконометрического моделирования

Вопросы для самоконтроля

1. Что измеряет эконометрика?
2. Каковы основныецелиэконометрики?
3. В чем состоят предмет и задачи эконометрики?
4. Каковы типы моделей и переменных, применяемых в эконометрике?
5. В чем особенности перекрестных и панельных данных?
6. В чем особенности временных рядов?
7. Что понимается под спецификацией модели?
8. Что такое параметризация?
9. Что понимается под верификацией модели?
10. В чем основное отличие эконометрической модели от математической?

Лекция 1(2)

Тема 3. Линейная модель парной регрессии и метод наименьших квадратов

Вопросы для изучения

1. Спецификация линейной модели парной регрессии.

2. Метод наименьших квадратов (МНК) – идентификация линейной модели парной регрессии.

3. Предпосылки МНК и свойства МНК-оценок.

Аннотация. Данная тема раскрывает суть регрессионного анализа в эконометрике.

Ключевые слова. Модель регрессии, метод наименьших квадратов, остатки регрессии.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где даются общие положения по теме.
- В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы.
- В качестве самостоятельной работы предлагается выполнить практические задания.
- Для проверки усвоения темы имеется тест для самоконтроля.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=382

2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none) С. 28-46.

3.Уткин, В. Б. Эконометрика [Электронный ресурс] : Учебник / В. Б. Уткин; Под ред. проф. В. Б. Уткина. - 2-е изд. - М.: Издательско-торговая корпорация «Дашков и К°», 2012. - 564 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 323-338.

4. Валентинов, В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 38-99.

5.Эконометрика. Практикум: [Электронный ресурс] Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов.знание, 2014. - 329 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 82-99.

6. Электронный курс “Econometrics and Public Policy (Advanced)”, Princeton University, URL: https://blackboard.princeton.edu/webapps/portal/frameset.jsp?tab_group=courses&url=%2Fwebapps%2Fblackboard%2Fexecute%2FcourseMain%3Fcourse_id%3D_214206_1

Спецификация линейной модели парной регрессии. 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:

$$y = \hat{f}(x) \qquad (1)$$

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

Множественная регрессия представляет собой регрессию результативного признака с двумя и большим числом факторов, т.е. модель вида:

$$y = \hat{f}(x_1, x_2, \ldots x_k) \qquad (2)$$

Любое эконометрическое исследование начинается со спецификации модели, т.е. с формулировки вида модели, исходя из соответствующей теории связи между переменными. Из всего круга факторов, влияющих на результативный признак, необходимо выделить наиболее существенно влияющие факторы. Парная регрессия достаточна, если имеется доминирующий фактор, который и используется в качестве объясняющей переменной. Например, выдвигается гипотеза о том, что величина спроса y на товар находится в обратной зависимости от цены x, т.е. $\hat{y}_x = a - b \cdot x.$

Уравнение простой регрессии характеризует связь между двумя переменными, которая проявляется как закономерность лишь в среднем по совокупности наблюдений. (Например, если зависимость спроса «y» от цены «x» имеет вид: y=5000-2x. Это означает, что с ростом цены на 1 д.е. спрос в среднем уменьшается на 2 д.е.). В уравнении регрессии корреляционная по сути связь признаков представляется в виде функциональной связи. В каждом отдельном случае величина y складывается из двух слагаемых:

$$y_j = \hat{y}_{x_j} + \varepsilon_j,$$

где $y_j -$ фактическое значение результативного признака; $\hat{y}_{x_j}$ - значение признака, найденное из математической функции связи y и x, т.е. из уравнения регрессии; ε_j - случайная величина, характеризующая отклонение реального значения признака от найденного по уравнению регрессии.

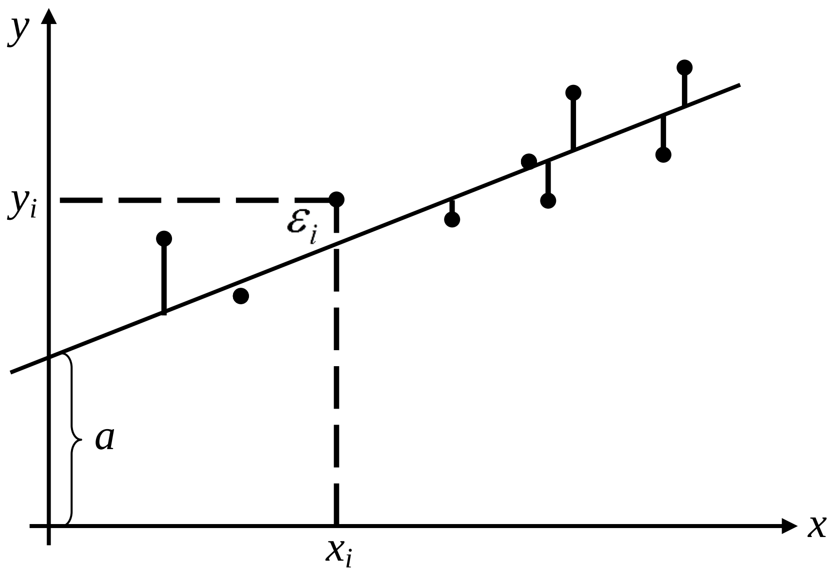
Случайная величина ε называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее порождают 3 источника: спецификация модели, выборочный характер исходных данных и ошибки измерения. Например, зависимость спроса от цены точнее следует записывать так: y=5000-2x+ε. В данном случае слева записано просто y, что означает фактическое значение, а не $\hat{y}$, отвечающее значению, рассчитанному по уравнению регрессии.

Метод наименьших квадратов (МНК) – идентификация линейной модели парной регрессии. Линейная регрессия сводится к нахождению уравнения вида

$$\hat{y}_x = a + bx \text{ (или } y = a + bx + \varepsilon\text{)} \qquad (3)$$

Первое выражение позволяет по заданным значениям фактора *x* рассчитать теоретические значения результативного признака, подставляя в него фактические значения фактора *x*. На графике теоретические значения лежат на прямой, которая представляют собой линию регрессии.

Построение линейной регрессии сводится к оценке ее параметров- *a* и *b*. Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК).



МНК позволяет получить такие оценки параметров *a* и *b*, при которых сумма квадратов отклонений фактических значений y от теоретических $\hat{y}_x$ минимальна:

$$\sum_i \left(y_i - \hat{y}_{x_i}\right)^2 \to \min, \text{ или } \sum_{i=1}^{n} \varepsilon_i^2 \to \min \qquad (4)$$

Для нахождения минимума надо вычислить частные производные суммы (4) по каждому из параметров - *a* и *b* - и приравнять их к нулю.

$$\Sigma\varepsilon_i^2 = S^2 = \Sigma(y - \hat{y}_x)^2 = \Sigma(y - a - bx)^2$$

$$\begin{cases} \dfrac{\partial s}{\partial a} = -2\sum y + 2na + 2\bar{b}\sum x = 0; \\ \dfrac{\partial s}{\partial b} = -2\sum y \cdot x + 2a\sum x + 2b\sum x^2 = 0 \end{cases} \qquad (5)$$

Преобразуем, получаем систему нормальных уравнений:

$$\begin{cases} n \cdot a + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum yx \end{cases} \qquad (6)$$

В этой системе n- объем выборки, суммы легко рассчитываются из исходных данных. Решаем систему относительно a и b, получаем:

$$b = \frac{n\sum yx - \left(\sum y\right)\left(\sum x\right)}{n\sum x^2 - \left(\sum x\right)^2}, \qquad (7)$$

$$a = \frac{1}{n}\sum y - \frac{b}{n}\sum x. \qquad (8)$$

Выражение (7) можно записать в другом виде:

$$b = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{cov(x, y)}{\sigma_x^2}, \qquad (9)$$

где $cov(x, y)$ –ковариация признаков, σ_x^2 –дисперсия фактора x.

Параметр b называется коэффициентом регрессии. Его величина показывает среднее изменение результата с изменением фактора на одну единицу. Возможность четкой экономической интерпретации коэффициента регрессии сделала линейное уравнение регрессии достаточно распространенным в эконометрических исследованиях.

Формально a- значение y при $x=0$. Если x не имеет и не может иметь нулевого значения, то такая трактовка свободного члена a не имеет смысла. Параметр a может не иметь экономического содержания. Попытки экономически интерпретировать его могут привести к абсурду, особенно при $a< 0$. Интерпретировать можно лишь знак при параметре a. Если $a> 0$, то относительное изменение результата происходит медленнее, чем изменение фактора. Сравним эти относительные изменения:

$bx < a+bx$ при $a>0,\ x>0 \Rightarrow b<\dfrac{a+bx}{x}$

$$\Rightarrow \frac{bdx}{dx.} < \frac{a+bx}{x} \Rightarrow \frac{dy}{dx} < \frac{y}{x} \Rightarrow \frac{dy}{y} < \frac{dx}{x}.$$

Иногда линейное уравнение парной регрессии записывают для отклонений от средних значений:

$$y' = b \cdot x', \tag{10}$$

где $y' = y - \bar{y}$, $x' = x - \bar{x}$. При этом свободный член равен нулю, что и отражено в выражении (10). Этот факт следует из геометрических соображений: уравнению регрессии отвечает та же прямая (3), но при оценке регрессии в отклонениях начало координат перемещается в точку с координатами $(\bar{x}, \bar{y})$. При этом в выражении (8) обе суммы будут равны нулю, что и повлечет равенство нулю свободного члена.

Предпосылки МНК и свойства МНК-оценок. Как было сказано выше, связь между y и x в парной регрессии является не функциональной, а корреляционной. Поэтому оценки параметров a и b являются случайными величинами, свойства которых существенно зависят от свойств случайной составляющей ε. Для получения по МНК наилучших результатов необходимо выполнение следующих предпосылок относительно случайного отклонения (условия Гаусса – Маркова):

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

2^0. Дисперсия случайных отклонений постоянна: $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$.

Выполнимость данной предпосылки называется гомоскедастичностью (постоянством дисперсии отклонений). Невыполнимость данной предпосылки называется гетероскедастичностью (непостоянством дисперсии отклонений)

3^0. Случайные отклонения ε_i и ε_j являются независимыми друг от друга для $i \neq j$:

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma^2, & i = j. \end{cases}$$

Выполнимость этого условия называется отсутствием автокорреляции.

4^0. Случайное отклонение должно быть независимо от объясняющих переменных. Обычно это условие выполняется автоматически, если объясняющие переменные в данной модели не являются случайными. Кроме того, выполнимость данной предпосылки для эконометрических моделей не столь критична по сравнению с первыми тремя.

При выполнимости указанных предпосылок имеет место теорема Гаусса-Маркова: оценки (7) и (8), полученные по МНК, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Таким образом, при выполнении условий Гаусса-Маркова оценки (7) и (8) являются не только несмещенными оценками коэффициентов регрессии, но и наиболее эффективными, т.е. имеют наименьшую дисперсию по сравнению с любыми другими оценками данных параметров, линейными относительно величин y_i.

Вопросы и задания для самоконтроля

1. Что такое функция регрессии?
2. Чем регрессионная модель отличается от функции регрессии?
3. Каковы основные причины наличия в регрессионной модели случайного отклонения?

4. Как осуществляется спецификация модели?

5. В чем состоит различие между теоретическим и эмпирическим уравнениями регрессии?

6. В чем суть метода наименьших квадратов?

7. Каковы формулы расчета коэффициентов эмпирического парного линейного уравнения регрессии по МНК?

8. Каковы предпосылки МНК? Каковы последствия их выполнимости или невыполнимости?

9. Действительно ли оценки коэффициентов регрессии будут иметь нормальное распределение, если случайные отклонения распределены нормально?

10. Действительно ли в любой линейной регрессионной модели, построенной по МНК, сумма случайных отклонений равна нулю?

Задача 1. При исследовании корреляционной зависимости между ценой на нефть X и индексом нефтяных компаний Y получены следующие данные: $\bar{x} = 16{,}2$; $\bar{y} = 4000$; $\sigma_x^2 = 4$; $\operatorname{cov}(x, y) = 40$.

Задание: построить линейное уравнение регрессии Y на X.

Задача 2. По выборке объема $n = 10$ получены следующие данные:

$$\sum x_i = 100;\ \sum y_i = 200;\ \sum x_i y_i = 21000;\ \sum x_i^2 = 12000;\ \sum y_i^2 = 45000.$$

Задание: оценить с помощью МНК параметры линейного уравнения регрессии, найти выборочный коэффициент корреляции r_{xy} .

Лекция 1(3)

Тема 4. Экономическая и статистическая интерпретация линейной модели парной регрессии

Вопросы для изучения:

1. Экономическая интерпретация параметров модели.

2. Коэффициенты корреляции и детерминации в линейной модели парной регрессии.

3. Проверка качества линейной модели парной регрессии (верификация модели).

4. Интервалы прогноза по линейному уравнению регрессии.

Аннотация. Данная тема раскрывает прикладное содержание регрессионного анализа.

Ключевые слова. Коэффициент регрессии, статистическая значимость, метод наименьших квадратов, коэффициент детерминации.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где даются общие представления по теме.
- В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы.
- В качестве самостоятельной работы предлагается выполнить практические задания.
- Для проверки усвоения темы имеется тест.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11766

2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none) С. 46-56.

3. Уткин, В. Б. Эконометрика [Электронный ресурс] : Учебник / В. Б. Уткин; Под ред. проф. В. Б. Уткина. - 2-е изд. - М.: Издательско-торговая корпорация «Дашков и К°», 2012. - 564 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 338-365.

4. Валентинов, В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 67-99.

5.Эконометрика. Практикум: [Электронный ресурс] Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов.знание, 2014. - 329 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 99-133.

Экономическая интерпретация параметров модели

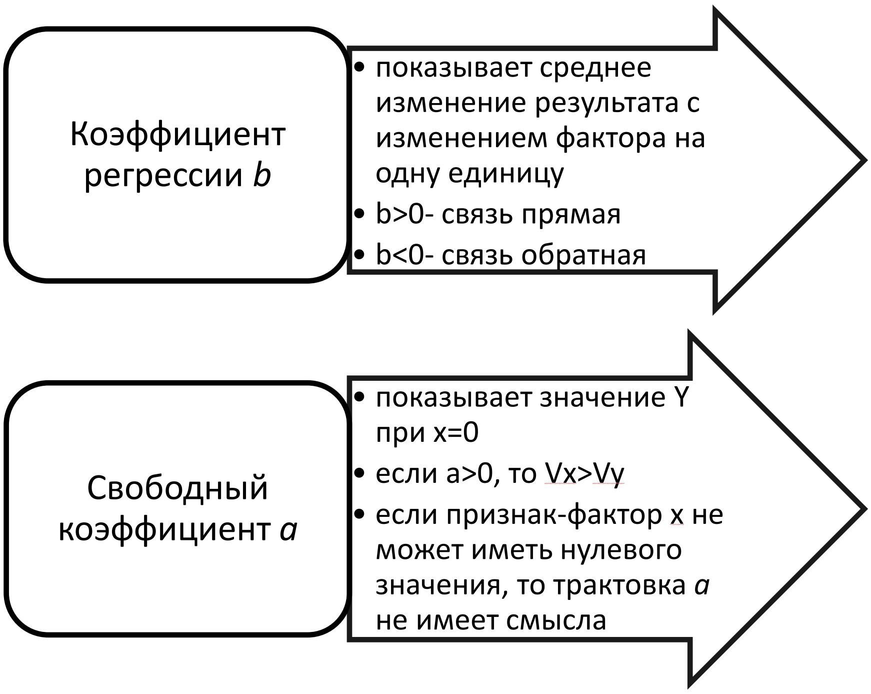


Рис. 4.1. Интерпретация параметров модели

Коэффициенты корреляции и детерминации в линейной модели парной регрессии. Если все точки лежат на построенной прямой, то регрессия Y на X «идеально» объясняет поведение зависимой переменной. Обычно поведение Y лишь частично объясняется влиянием переменной X.

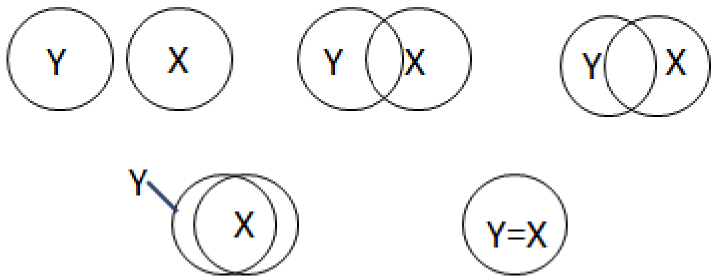


Рис. 4.2. Диаграмма Венна

Линейный коэффициент парной корреляции:

$$r_{yx} = b\frac{\sigma_x}{\sigma_y} = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y}\cdot\bar{x}}{\sigma_x \cdot \sigma_y}$$

$$-1 \le r_{yx} \le 1$$

Если b>0, то $r_{yx}>0$; если b<0, то $r_{yx}<0$.

По абсолютной величине, чем ближе значение r_{xy} к единице, тем теснее связь, чем ближе значение r_{xy} к нулю, тем слабее связь между *y* и *x*.

$$|r_{yx}| < 0{,}3 - \textit{слабая}$$

$$0{,}3 \le |r_{yx}| \le 0{,}7 - \textit{средняя}$$

$$|r_{yx}| > 0{,}7 - \textit{сильная}, \textit{тесная}$$

Суммы квадратов отклонений:

- общая (TSS): $\sum (y_i - \bar{y})^2$

- регрессионная (ESS): $\sum (y_x - \bar{y})^2$

- остаточная (RSS): $\sum (y_i - y_x)^2$

$$\sum (y_i - \bar{y})^2 = \sum (y_x - \bar{y})^2 + \sum (y_i - y_x)^2$$

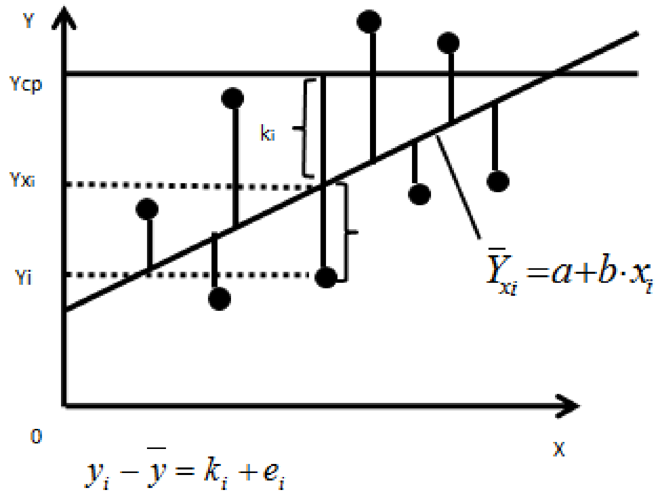


Рис.4.3. Геометрическая интерпретация

Число степеней свободы (df-degrees of freedom) - это число независимо варьируемых значений признака. Для общей СКО требуется *(n-1)* независимых отклонений, т.к. $\sum(y - \bar{y}) = 0,$ что позволяет свободно варьировать *(n-1)* значений, а последнее n-е отклонение определяется из общей суммы, равной нулю. Поэтому $df_{общ.} = n - 1.$

Факторную СКО можно выразить так:

$$\sum(\hat{y}_x - \bar{y})^2 = \sum[(a + bx) - (a + b\bar{x})]^2$$
$$= \sum(bx - b\bar{x})^2 = b^2\sum(x - \bar{x})^2$$

Эта СКО зависит только от одного параметра - *b*,поскольку выражение под знаком суммы к значениям результативного признака не относится. Следовательно, факторная СКО имеет одну степень свободы, и $df_{факт.} = 1.$

Для определения $df_{\text{остат.}}$ воспользуемся аналогией с балансовым равенством (11). Можно записать равенство и между числами степеней свободы:

$$df_{\text{общ.}} = df_{\text{факт.}} + df_{\text{остат.}}$$

Таким образом, можем записать: $(n-1) = 1 + (n-2)$

Из этого баланса определяем, что $df_{\text{остат.}}$=*n-2*.

Разделив каждую СКО на свое число степеней свободы, получим средний квадрат отклонений, или дисперсию на одну степень свободы.

Выборочные оценки дисперсий:

- общая дисперсия: $S^2_{TSS} = \dfrac{\sum(y_i - \overline{y})^2}{n-1}$

- регрессионная дисперсия: $S^2_{ESS} = \dfrac{\sum(y_x - \overline{y})^2}{m}$

- остаточная дисперсия: $S^2_{RSS} = \dfrac{\sum(y_i - y_x)^2}{n-m-1}$

Коэффициент детерминации:

$$R^2 = \frac{\sum(y_x - \overline{y})^2}{\sum(y_i - \overline{y})^2} = 1 - \frac{\sum(y_i - y_x)^2}{\sum(y_i - \overline{y})^2}$$
$$R^2 = r^2_{yx}; 0 \le R^2 \le 1$$

Коэффициент детерминации определяет долю разброса зависимой переменной Y, объясняемую регрессией Y на X.

Проверка качества модели линейной парной регрессии (верификация модели)

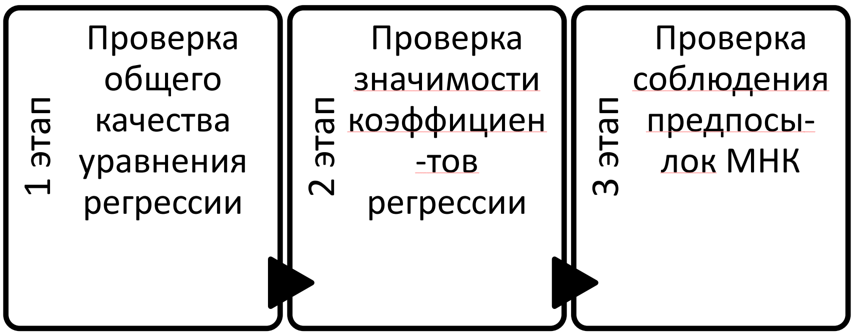


Рис. 4.5. Этапы проверки качества модели

1 этап: F-тест состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи.

$$H_0 : D^2{}_{ESS} = D^2{}_{RSS}$$
$$H_1 : D^2{}_{ESS} > D^2{}_{RSS}$$
$$F = \frac{\sum (y_x - \bar{y})^2 / m}{\sum (y_i - y_x)^2 /(n-m-1)} = \frac{r^2{}_{xy}}{1-r^2{}_{xy}} \cdot (n-2)$$
$$F > F_{\alpha, v_1 = m, v_2 = n-m-1} \Rightarrow H_1$$
$$F < F_{\alpha, v_1 = m, v_2 = n-m-1} \Rightarrow H_0$$

2 этап: T-тест состоит в проверке гипотезы H_0 о статистической незначимости коэффициентов регрессии и корреляции.

$$H_0 : \beta = 0$$
$$H_1 : \beta \neq 0$$
$$|t_b| > t_{\alpha/2, n-2} \Rightarrow H_1$$
$$|t_b| < t_{\alpha/2, n-2} \Rightarrow H_0$$

$$_b = \frac{b}{m_b}; t_a = \frac{a}{m_a}; t_r = \frac{r}{m_r} = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}$$
$$m_b = \sqrt{\frac{\sum (y - y_x)^2 /(n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2{}_{RSS}}{\sum (x - \bar{x})^2}} = \frac{S_{RSS}}{\sigma_x \sqrt{n}}$$

$$m_a = \sqrt{\frac{\sum (y - y_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}}$$
$$m_r = \sqrt{\frac{1-r^2}{n-2}}; t^2{}_r = t^2{}_b = F$$

3 этап: проведение тестов на гетероскедастичность и автокорреляцию остатков.

Доверительные интервалы для коэффициентов теоретического уравнения регрессии:

$$t = \frac{b - \beta}{m_b}$$
$$\Delta b = t_{\alpha/2, n-2} \cdot m_b;$$
$$b - \Delta b < \beta < b + \Delta b$$
$$\Delta a = t_{\alpha/2, n-2} \cdot m_a;$$
$$a - \Delta a < \alpha < a + \Delta a$$

Интервалы прогноза по линейному уравнению регрессии. Прогнозирование по уравнению регрессии представляет собой подстановку в уравнение регрессии соответственного значения *x*. Такой прогноз $\hat{y}_x$ называется точечным. Он не является точным, поэтому дополняется расчетом стандартной ошибки $\hat{y}_x$; получается интервальная оценка прогнозного значения y^*:

$$\hat{y}_x - m_{\hat{y}_x} \le y^* \le \hat{y} + m_{\hat{y}_x}$$

Преобразуем уравнение регрессии:

$$\hat{y}_x = a + bx = (\bar{y} - b\bar{x}) + bx = \bar{y} + b(x - \bar{x})$$

ошибка $m_{\hat{y}_x}$ зависит от ошибки $\bar{y}$ и ошибки коэффициента регрессии b, т.е. $m^2_{\hat{y}_x} = m^2_{\bar{y}} + m^2_b (x - \bar{x})^2.$

Из теории выборки известно, что $m^2_{\bar{y}} = \dfrac{\sigma^2}{n}$

Используем в качестве оценки σ^2 остаточную дисперсию на одну степень свободы S^2, получаем: $m^2_{\bar{y}} = \dfrac{S^2}{n}$

Ошибка коэффициента регрессии :

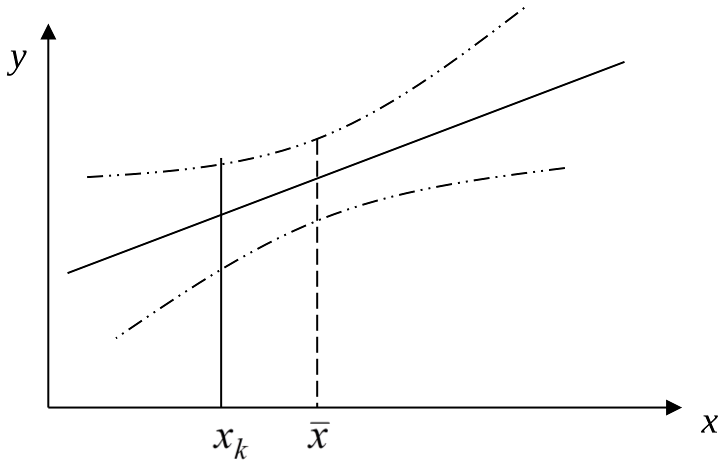
$$m^2_b = \frac{S^2}{\sum (x - \bar{x})^2}$$

Таким образом, при $x = x_к$ получаем:

$$m_{\hat{y}_x}^2 = \frac{S^2}{n} + \frac{S^2}{\sum (x-\bar{x})^2}(x_{\hat{e}} - \bar{\tilde{o}})^2 = S^2\left(\frac{1}{n} + \frac{(x_{\hat{e}} - \bar{\tilde{o}})^2}{\sum (\tilde{o} - \bar{\tilde{o}})^2}\right)$$

$$m_{\hat{y}_x} = S\sqrt{\frac{1}{n} + \frac{(x_{\hat{e}} - \bar{\tilde{o}})^2}{\sum (\tilde{o} - \bar{\tilde{o}})^2}} \qquad (11)$$

Как видно из формулы (11), величина $m_{\hat{y}_x}$ достигает минимума при $x_к = \bar{x}$ и возрастает по мере удаления $x_к$ от $\bar{x}$ в любом направлении.



Для нашего примера эта величина составит:

$$m_{\hat{y}_x} = \sqrt{53\left(\frac{1}{7} + \frac{(x_{\hat{e}} - 3{,}143)^2}{10{,}857}\right)}$$

При $x_к = \bar{x}$ $m_{\hat{y}_x} = \sqrt{53:7} = 2{,}75$. При $x_к = 4$

$$m_{\hat{y}_x} = \sqrt{53\left(\frac{1}{7} + \frac{(4 - 3{,}143)^2}{10{,}857}\right)} = 3{,}34$$

Для прогнозируемого значения $\tilde{o}_{\tilde{o}}$ 95% - ные доверительные интервалы при заданном $x_к$ определены выражением:

$$\tilde{o}_{\tilde{o}_{\hat{e}}} \pm t_\alpha \cdot m_{\hat{y}_x,} \qquad (12)$$

т.е. при $x_{к} = 4$ $\hat{y}_{x_{ê}} \pm 2{,}57 \cdot 3{,}34$ или $\hat{y}_{x_{ê}} \pm 8{,}58$. При $x_{к} = 4$ прогнозное значение составит $y_p = -5{,}79 + 36{,}84 \cdot 4 = 141{,}57$ - это точечный прогноз.

Прогноз линии регрессии лежит в интервале:

$132{,}99 \le \tilde{o}_{\tilde{o}_{ê}} \le 150{,}15$

Мы рассмотрели доверительные интервалы для среднего значения y при заданном x. Однако фактические значения y варьируются около среднего значения $\tilde{o}_{\tilde{o}}$, они могут отклоняться на величину случайной ошибки ε, дисперсия которой оценивается как остаточная дисперсия на одну степень свободы S^2. Поэтому ошибка прогноза отдельного значения y должна включать не только стандартную ошибку $m_{\hat{y}_x}$, но и случайную ошибку *S*. Таким образом, средняя ошибка прогноза индивидуального значения y составит:

$$m_{y_{i(x_к)}} = S\sqrt{1 + \frac{1}{n} + \frac{(x_к - \bar{x})^2}{\sum (x - \bar{x})^2}} \qquad (13)$$

Для примера:

$$m_{y_{i(x_к=4)}} = \sqrt{53 \cdot \left(1 + \frac{1}{7} + \frac{(4 - 3{,}143)^2}{10{,}857}\right)} = 8{,}01$$

Доверительный интервал прогноза индивидуальных значений y при $x_к = 4$ с вероятностью 0,95 составит: $141{,}57 \pm 2{,}57 \cdot 8{,}01$, или $120{,}98 \le y_p \le 162{,}16$.

Пусть в примере с функцией издержек выдвигается предположение, что в предстоящем году в связи со стабилизацией экономики затраты на производство 8 тыс. ед. продукции не превысят 250 млн. руб. Означает ли это изменение найденной закономерности или затраты соответствуют регрессионной модели?

Точечный прогноз: $\hat{y}_{x=8} = -5{,}79 + 36{,}84 \cdot 8 = 288{,}93.$

Предполагаемое значение - 250. Средняя ошибка прогнозного индивидуального значения:

$$m_{y_{i(x_i)}} = S\sqrt{1 + \frac{1}{n} + \frac{(x_K - \bar{x})}{\sum(x - \bar{x})}} =$$

$$= \sqrt{53 \cdot \left(1 + \frac{1}{7} + \frac{(8 - 3{,}143)^2}{10{,}857}\right)} = 13{,}26$$

Сравним ее с предполагаемым снижением издержек производства, т.е. 250-288,93=-38,93:

$$t = \frac{-38{,}93}{12{,}26} = -2{,}93.$$

Поскольку оценивается только значимость уменьшения затрат, то используется односторонний t- критерий Стьюдента. При ошибке в 5 % с $n-2=5$ $\quad t_{таб.} = 2{,}015$, поэтому предполагаемое уменьшение затрат значимо отличается от прогнозируемого значения при 95 % - ном уровне доверия. Однако, если увеличить вероятность до 99%, при ошибке 1 % фактическое значение t – критерия оказывается ниже табличного 3,365, и различие в затратах статистически не значимо, т.е. затраты соответствуют предложенной регрессионной модели.

Вопросы и задания для самоконтроля

1. Каков экономический смысл коэффициента регрессии?
2. Какой смысл может иметь свободный коэффициент уравнения регрессии?
3. Какова связь между линейным коэффициентом корреляции и коэффициентом регрессии в линейной модели парной регрессии?
4. Каков статистический смысл коэффициента детерминации?

5. Как записывается баланс для сумм квадратов отклонений результативного признака?

6. Что происходит, когда общая СКО равна остаточной? В каком случае общая СКО равна факторной?

7. Что такое число степеней свободы? Чему равны числа степеней свободы для различных СКО в парной регрессии?

8. Как используется F-статистика в регрессионном анализе?

9. Как F-статистика связана с коэффициентом детерминации в парной регрессии?

10. Как рассчитать критерий Стьюдента для коэффициента регрессии в линейной модели парной регрессии?

11. В чем суть предсказания индивидуальных значений зависимой переменной?

Задача 1. Пусть имеется следующая модель парной регрессии, построенная по 20 наблюдениям: $\tilde{y} = 8 - 7x$. При этом r_{xy} = - 0,5.

Задание: построить доверительный интервал для коэффициента регрессии в этой модели с вероятностями 0,9 и 0,95.

Задача 2. Анализируется зависимость между доходами горожан (X), имеющими индивидуальные домовладения, и рыночной стоимостью их домов (Y). По случайной выборке из 120 горожан данной категории получены результаты:

$$\sum x_i = 27343;\ \sum y_i = 115870;\ \sum (x_i - \bar{x})^2 = 75200;$$

$$\sum (y_i - \bar{y})^2 = 1620340;\ \sum (x_i - \bar{x})(y_i - \bar{y}) = 250431.$$

Задание: найти оценку коэффициента регрессии b_1 и построить 95% доверительный интервал для коэффициента регрессии.

Лекция 2(1)

Тема 5. Линейная модель множественной регрессии, оценка ее параметров

Вопросы для изучения

1. Линейная модель множественной регрессии. Эмпирическая форма записи. Оценка параметров модели с помощью МНК.
2. Показатели качества множественной регрессии.
3. Мультиколлинеарность.

Аннотация. Данная тема раскрывает особенности линейной модели множественной регрессии.

Ключевые слова. Стандартизованный коэффициент регрессии, метод наименьших квадратов, мультиколлинеарность.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где даются общие представления по теме.
- В дополнение к лекции есть презентация, которую необходимо изучить и ответить на вопросы.
- В качестве самостоятельной работы предлагается выполнить практические задания.
- Для проверки усвоения темы имеется тест.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11766
2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none) С.

3. Уткин, В. Б. Эконометрика [Электронный ресурс] : Учебник / В. Б. Уткин; Под ред. проф. В. Б. Уткина. - 2-е изд. - М.: Издательско-торговая корпорация «Дашков и К°», 2012. - 564 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none)

4. Валентинов, В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 142-181.

5.Эконометрика. Практикум: [Электронный ресурс] Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов.знание, 2014. - 329 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 133-140.

6. Электронный курс “Econometrics and Public Policy (Advanced)”, Princeton University, URL: https://blackboard.princeton.edu/webapps/portal/frameset.jsp?tab_group=courses&url=%2Fwebapps%2Fblackboard%2Fexecute%2FcourseMain%3Fcourse_id%3D_214206_1

Линейная модель множественной регрессии. Эмпирическая форма записи. Оценка параметров модели с помощью МНК. На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов. Например, спрос на некоторое благо определяется не только ценой данного блага, но и ценами на замещающие и дополняющие блага, доходом потребителей и многими другими факторами. В этом случае вместо парной регрессии рассматривается множественная регрессия $\hat{y} = f(x_1, x_2, ..., x_p)$.

Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства, в макроэкономических расчетах и в ряде других вопросов экономики. В настоящее время множественная регрессия – один из наиболее распространенных методов в эконометрике. Основной целью множественной регрессии является построение модели с большим числом факторов, а также определение влияния каждого фактора в отдельности и совокупного их воздействия на моделируемый показатель.

Множественный регрессионный анализ является развитием парного регрессионного анализа в случаях, когда зависимая переменная связана более чем с одной независимой переменной. Большая часть анализа является непосредственным расширением парной регрессионной модели, но здесь также появляются и некоторые новые проблемы, из которых следует выделить две. Первая проблема касается исследования влияния конкретной независимой переменной на зависимую переменную, а также разграничения её воздействия и воздействий других независимых переменных. Второй важной проблемой является спецификация модели, которая состоит в том, что необходимо ответить на вопрос, какие факторы следует включить в регрессию (1), а какие – исключить из неё. В дальнейшем изложение общих вопросов множественного регрессионного анализа будем вести, разграничивая эти проблемы. Поэтому вначале будем полагать, что спецификация модели правильна.

Самой употребляемой и наиболее простой из моделей множественной регрессии является линейная модель множественной регрессии:

$$y = \alpha' + \beta_1' x_1 + \beta_2' x_2 + \ldots + \beta_p' x_p + \varepsilon \qquad (1)$$

По математическому смыслу коэффициенты β_j' в уравнении (1) равны частным производным результативного признака y по соответствующим факторам:

$$\beta_1' = \frac{\partial y}{\partial x_1}, \beta_2' = \frac{\partial y}{\partial x_2}, \ldots, \beta_p' = \frac{\partial y}{\partial x_p}.$$

Параметр α называется свободным членом и определяет значение y в случае, когда все объясняющие переменные равны нулю. Однако, как и в случае парной регрессии, факторы по своему экономическому содержанию часто не могут принимать нулевых значений, и значение свободного члена не имеет экономического смысла. При этом, в отличие от парной регрессии, значение каждого регрессионного коэффициента β'_j равно среднему изменению y при увеличении x_j на одну единицу лишь при условии, что все остальные факторы остались неизменными. Величина ε представляет собой случайную ошибку регрессионной зависимости. Поскольку параметры $\alpha', \beta_1', \beta_2'..., \beta_p'$ являются случайными величинами, определить их истинные значения по выборке невозможно. Поэтому вместо теоретического уравнения регрессии оценивается так называемое эмпирическое уравнение множественной регрессии, которое можно представить в виде:

$$y = a + b_1 x_1 + b_2 x_2 + ... + b_p x_p + e \qquad (2)$$

Здесь $a, b_1, b_2, ..., b_p$ - оценки теоретических значений $\alpha', \beta_1', \beta_2'..., \beta_p'$, или эмпирические коэффициенты регрессии, e – оценка отклонения ε. Тогда расчетное выражение имеет вид: $\hat{y} = a + b_1 x_1 + b_2 x_2 + ... + b_p x_p$

Пусть имеется n наблюдений объясняющих переменных и соответствующих им значений результативного признака:

$$\left(x_{i1}, x_{i2}, ..., x_{ip}, y_i\right), \quad i = \overline{1, n}$$

Для однозначного определения значений параметров эмпирического уравнения множественной регрессии объем выборки n должен быть не меньше количества параметров, т.е. $n \geq p + 1$. В противном случае значения параметров не могут быть определены однозначно. Для получения надежных оценок параметров уравнения объём выборки должен значительно превышать количество определяемых по нему параметров. Практически, как было сказано ранее, объём выборки должен превышать количество параметров при x_j в уравнении (4) в 6-7 раз.

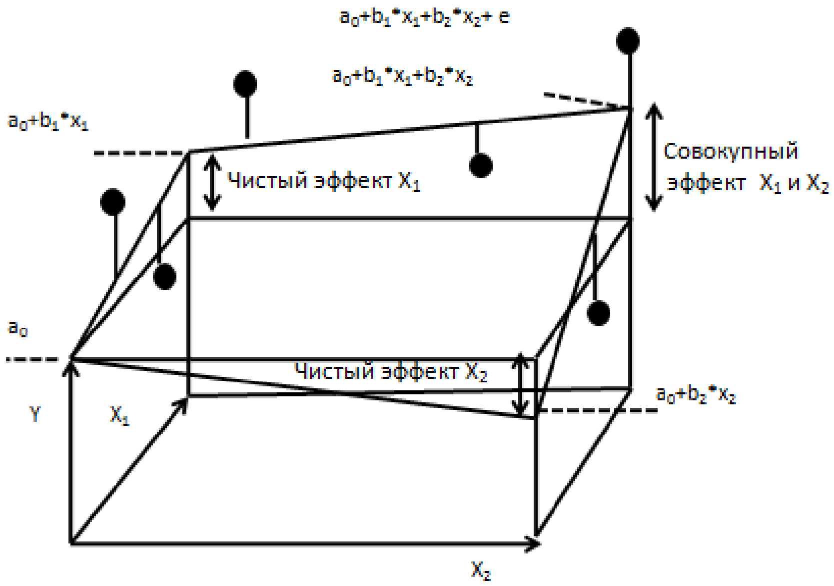


Рис.5.1. Геометрическая интерпретация линейной модели множественной регрессии

$Y = a_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + e$, где

Y-общая величина расходов на питание;

X_1- располагаемый личный доход;

X_2- цена продуктов питания.

Экономическая интерпретация: При каждом увеличении располагаемого личного дохода X_1 на 1 единицу собственного измерения, расходы на питание (Y) увеличиваются на b_1 единиц измерения при сохранении постоянных цен. На каждую единицу индекса цен X_2 эти расходы уменьшаются на b_2 единиц измерения при сохранении постоянных доходов. Если $a_0>0$, то вариация расходов меньше вариации факторов; если $a_0<0$, то вариация расходов больше вариации факторов.

Для проведения анализа в рамках линейной модели множественной регрессии необходимо выполнение ряда предпосылок МНК. В основном это те же предпосылки, что и для парной регрессии, однако здесь нужно добавить предположения, специфичные для множественной регрессии:

5^0.Спецификация модели имеет вид: $y = \alpha' + \beta_1' x_1 + \beta_2' x_2 + ... + \beta_p' x_p + \varepsilon$.

6^0.Отсутствие мультиколлинеарности: между объясняющими переменными: отсутствует строгая линейная зависимость, что играет важную роль в отборе факторов при решении проблемы спецификации модели.

7^0.Ошибки $\varepsilon_i, i=\overline{1,n}$ имеют нормальное распределение $(\varepsilon_i \sim N(0,\sigma))$. Выполнимость этого условия нужна для проверки статистических гипотез и построения интервальных оценок.

При выполнимости всех этих предпосылок имеет место многомерный аналог теоремы Гаусса – Маркова: оценки $a, b_1, b_2, ..., b_p$, полученные по МНК, являются наиболее эффективными (в смысле наименьшей дисперсии) в классе линейных несмещенных оценок.

2 вопроса спецификации модели	
Отбор факторов	Выбор вида уравнения регрессии

Рис.5.2. Проблемы спецификации модели

Кроме этого, факторы, включаемые во множественную регрессию, должны быть количественно измеримы.

Рассмотрим три метода расчета параметров множественной линейной регрессии.

1. *Матричный метод.* Представим данные наблюдений и параметры модели в матричной форме.

$Y=[y_1, y_2, ..., y_n]'$ - n – мерный вектор – столбец наблюдений зависимой переменной;

$B=[a, b_1, b_2, ..., b_p]'$ - (p+1) – мерный вектор – столбец параметров уравнения регрессии $y=a+b_1x_1+b_2x_2+...+b_px_p+e$;

$Y=[y_1,y_2,...,y_n]'$ - n – мерный вектор – столбец отклонений выборочных значений y_i от значений $\hat{y}_i$, получаемых по уравнению $\hat{y}=a+b_1x_1+b_2x_2+...+b_px_p$.

Для удобства записи столбцы записаны как строки и поэтому снабжены штрихом для обозначения операции транспонирования.

Наконец, значения независимых переменных запишем в виде прямоугольной матрицы размерности $n\times(p+1)$:

$$X=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Каждому столбцу этой матрицы отвечает набор из n значений одного из факторов, а первый столбец состоит из единиц, которые соответствуют значениям переменной при свободном члене.

В этих обозначениях эмпирическое уравнение регрессии выглядит так: $Y=XB+e$. Отсюда вектор остатков регрессии можно выразить таким образом: $e=Y-XB$.

Таким образом, функционал $Q=\sum e_i^2$, который, собственно, и минимизируется по МНК, можно записать как произведение вектора – строки ***e'*** на вектор – столбец ***e***: $Q=e'e=(Y-XB)'(Y-XB)$.

В соответствии с МНК дифференцирование $\boldsymbol{Q}$ по вектору $\boldsymbol{B}$ приводит к выражению: $\frac{\partial Q}{\partial B}=-2X'Y+2(X'X)B$, которое для нахождения экстремума следует приравнять к нулю. В результате преобразований получаем выражение для вектора параметров регрессии: $B=(X'X)^{-1}X'Y$. Здесь $(X'X)^{-1}$ - матрица, обратная к $X'X$.

Пример. Бюджетное обследование пяти случайно выбранных семей дало следующие результаты (в тыс. руб.):

Семья	Накопления, ***S***	Доход, ***Y***	Имущество, ***W***
1	3	40	60
2	6	55	36
3	5	45	36
4	3,5	30	15
5	1,5	30	90

Оценим регрессию ***S*** на ***Y*** и ***W***. Введем обозначения:

S=[3;6;5;3,5;1,5]’ – вектор наблюдений зависимой переменной;

B=[$a;b_1;b_2$]’ – вектор параметров уравнения регрессии;

$$X=\begin{bmatrix}1 & 40 & 60\\ 1 & 55 & 36\\ 1 & 45 & 36\\ 1 & 30 & 15\\ 1 & 30 & 90\end{bmatrix}$$

- матрица значений независимых переменных.

Далее с помощью матричных операций вычисляем (используем табличный процессор MS Excel и функции ТРАНСП, МУМНОЖ и МОБР в нем):

$$X'X=\begin{bmatrix}5 & 200 & 237\\ 200 & 8450 & 9150\\ 237 & 9150 & 14517\end{bmatrix};\quad (X'X)^{-1}=\begin{bmatrix}5{,}6916 & -0{,}1074 & -0{,}0252\\ -0{,}1074 & 0{,}0024 & 0{,}00024\\ -0{,}0252 & 0{,}00024 & 0{,}00033\end{bmatrix}$$

$$B=(X'X)^{-1}X'Y=\begin{pmatrix}0{,}2787 & 0{,}1229 & -0{,}0294\end{pmatrix}$$

Регрессионная модель в скалярном виде:

$$\hat{S}=0{,}2787+0{,}1229Y-0{,}0294W$$

2. Скалярный метод. При его применении строится система нормальных уравнений, решение которой и позволяет получить оценки параметров регрессии:

$$\begin{cases} an & +b_1\sum x_1 & +b_2\sum x_2 & +\ldots+b_p\sum x_p & =\sum y \\ a\sum x_1 & +b_1\sum x_1^2 & +b_2\sum x_2x_1 & .+\ldots+b_p\sum x_px_1 & =\sum yx_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a\sum x_p & +b_1\sum x_1x_p & +b_2\sum x_2x_p & +\ldots+b_p\sum x_p^2 & =\sum yx_p \end{cases}$$

Решить эту систему можно любым подходящим способом, например, методом определителей или методом Гаусса. При небольшом количестве определяемых параметров использование определителей предпочтительнее.

Рассмотрим пример, приведенный выше. Здесь для двух факторов, *Y* и *W*, система нормальных уравнений запишется так:

$$\begin{cases} an & +b1\sum Y & +b2\sum W & =\sum S \\ a\sum Y & +b1\sum Y^2 & +b2\sum WY & =\sum SY \\ a\sum W & +b1\sum YW & +b2\sum W^2 & =\sum SW \end{cases}$$

Рассчитываем значения сумм, получаем:

$$\begin{cases} 5a & +200b1 & +237b2 & =19 \\ 200a & +8450b1 & +9150b2 & =825 \\ 237a & +9150b1 & +14517b2 & =863{,}5 \end{cases}$$

Рассчитаем значения определителей этой системы, используем функцию МОПРЕД в Excel:

$$\Delta = 6842700; \quad \Delta_a = 1903325; \quad \Delta_{b_1} = 840825; \quad \Delta_{b_2} = -201225.$$

Отсюда получим оценки параметров модели:

$$a = \Delta / \Delta_a = 1903325/6842700 = 0{,}2787;$$
$$b_1 = \Delta_{b_1} / \Delta = 840825/6842700 = 0{,}1229;$$
$$b_2 = \Delta_{b_2} / \Delta = -201205/6842700 = -0{,}0294.$$

Обратите внимание, что коэффициенты в левой части системы нормальных уравнений совпадают с соответствующими элементами матрицы $X'X$.

3. Регрессионная модель в стандартизованном масштабе. Уравнение регрессии в стандартизованном масштабе имеет вид:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_p t_{x_p} + \varepsilon$$

где $t_y, t_{x_1}, t_{x_2}, ..., t_{x_p}$ - стандартизованные переменные:

$$t_y = \frac{y - \bar{y}}{\sigma_y}; \quad t_{x_j} = \frac{x_j - \bar{x}_j}{\sigma_{x_j}}, \quad j = \overline{1, n}$$

для которых среднее значение равно нулю: $\bar{t}_y = \bar{t}_{x_1} = \bar{t}_{x_2} = ... = \bar{t}_{x_p} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_y = \sigma_{t_{x_j}} = 1,\ j = \overline{1, n}$; β_j – стандартизованные коэффициенты регрессии, или β – коэффициенты (не следует путать их с параметрами уравнения $y = \alpha' + \beta_1' x_1 + \beta_2' x_2 + ... + \beta_p' x_p + \varepsilon$).

Применяя МНК к уравнению $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_p t_{x_p} + \varepsilon$, после соответствующих преобразований получим систему нормальных уравнений:

$$\begin{cases} \beta_1 & + \beta_2 r_{x_2 x_1} & + \beta_3 r_{x_3 x_1} & + \beta_p r_{x_p x_1} & = r_{y x_1} \\ \beta_1 r_{x_1 x_2} & + \beta_2 & + \beta_3 r_{x_3 x_2} & + \beta_p r_{x_p x_2} & = r_{y x_2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \beta_1 r_{x_1 x_p} & + \beta_2 r_{x_2 x_p} & + \beta_3 r_{x_3 x_p} & + \beta_p & = r_{y x_p} \end{cases}$$

В этой системе $r_{y x_j}$, $r_{x_i x_j}$, $j, k = \overline{1, p}$ - элементы расширенной матрицы парных коэффициентов корреляции или, другими словами, коэффициенты парной корреляции между различными факторами или между факторами и результативным признаком. Имея измеренные значения всех переменных, вычислить матрицу парных коэффициентов корреляции на компьютере не составляет большого труда, используя, например, табличный процессор MS Excel или программу Statistica.

Решением данной системы определяются β – коэффициенты. Эти коэффициенты показывают, на сколько значений с.к.о. изменится в среднем результат, если соответствующий фактор x_j изменится на одну с.к.о. при неизменном среднем уровне других факторов. Поскольку все переменные заданы как центрированные и нормированные, β – коэффициенты сравнимы между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии, в отличие от коэффициентов обычной регрессии, которые несравнимы между собой.

Пусть функция издержек производства y (тыс. руб.) характеризуется уравнением вида: $y = 200 + 1{,}2x_1 + 1{,}1x_2 + \varepsilon$, где факторами являются основные производственные фонды (тыс. руб.) и численность занятых в производстве (чел.). Отсюда видно, что при постоянной занятости рост стоимости основных производственных фондов на 1 тыс. руб. влечет за собой увеличение затрат в среднем на 1,2 тыс. руб., а увеличение числа занятых на одного человека при неизменной технической оснащенности приводит к росту затрат в среднем на 1,1 тыс. руб.. Однако это не означает, что первый фактор сильнее влияет на издержки производства по сравнению со вторым. Такое сравнение возможно, если обратиться к уравнению регрессии в стандартизованном масштабе. Пусть оно выглядит так: $\hat{t}_y = 0{,}5t_{x_1} + 0{,}8t_{x_2}$. Это означает, что с ростом первого фактора на одно с.к.о. при неизменном числе занятых затраты на продукцию увеличиваются в среднем на 0,5 с.к.о. Так как $\beta_1 < \beta_2$ (0,5<0,8), то можно заключить, что большее влияние на производство продукции оказывает второй фактор, а не первый, как кажется из уравнения регрессии в натуральном масштабе.

В парной зависимости стандартизованный коэффициент регрессии есть не что иное, как линейный коэффициент корреляции r. Подобно тому, как в парной зависимости коэффициенты регрессии и корреляции связаны между собой, так и во множественной регрессии коэффициенты «чистой» регрессии b_j связаны с β – коэффициентами: $b_j = \beta_j \dfrac{\sigma_y}{\sigma_{x_j}}$.

Это позволяет от уравнения регрессии в стандартизованном масштабе: $\hat{t}_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_p t_{x_p}$ переходить к уравнению регрессии в натуральном масштабе $\hat{y} = a + b_1x_1 + b_2x_2 + ... + b_px_p$. Параметр a определяется так: $a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - ... - b_p\bar{x}_p$.

Свободный член в уравнении $\hat{t}_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_p t_{x_p}$ отсутствует, поскольку все стандартизованные переменные имеют нулевое среднее значение.

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

Компьютерные программы построения уравнения множественной регрессии в зависимости от использованного в них алгоритма решения позволяют получить либо только уравнение регрессии для исходных данных, либо, кроме того, уравнение регрессии в стандартизованном масштабе.

В заключение приведем расчет стандартизованного уравнения регрессии по данным рассмотренного выше числового примера. Используя функцию КОРРЕЛ в Excel, рассчитаем расширенную матрицу парных коэффициентов корреляции:

$$R = \begin{bmatrix} 1 & -0,27149 & 0,873684 \\ -0,27149 & 1 & -0,68224 \end{bmatrix},$$

в которой последний столбец состоит из элементов $r_{yx_1}(r_{SY})$ и $r_{SW}(r_{yx_2})$ соответственно, а неединичные элементы в первых двух столбцах соответствуют $r_{YW}(r_{x_1x_2})$. Эта матрица является расширенной матрицей системы уравнений для определения β – коэффициентов:

$$\begin{cases} \beta_1 + 0,27149\beta_2 = 0,873684, \\ -0,27149\beta_1 + \beta_2 = -0,68224 \end{cases}$$

Решаем систему методом определителей, получаем:

Δ=0,926291; Δ_1=0,688461; Δ_2=-0,44504;

β_1=0,688461/0,926291=0,743245;

β_2=-0,44504/0,926291=-0,48045;

Тогда стандартизованное уравнение регрессии запишется так:

$$\hat{t}_y = 0{,}743245 t_Y - 0{,}48045 t_W$$

Отсюда видно, что первый фактор оказывает большее воздействие на результат, чем второй ($|\beta_1|>|\beta_2|$), однако эта разница не так велика, как для коэффициентов в натуральном масштабе (0,1229 и –0,0294). От этого уравнения можно перейти к уравнению в натуральном масштабе. Для этого с помощью функции СТАНДОТКЛОН в Excel определим стандартные отклонения всех переменных: $\sigma_S = 1{,}75357;\quad \sigma_Y = 10{,}6066;\quad \sigma_W = 28{,}6496,$

а с помощью функции СРЗНАЧ – средние значения: $\bar{S} = 3{,}8;\quad \bar{Y} = 40;\quad \bar{W} = 47{,}4.$

Далее определяем оценки параметров:

$$b_1 = \beta_1 \frac{\sigma_y}{\sigma_{x_1}} = 0{,}743245 \cdot \frac{1{,}75357}{10{,}6066} = 0{,}1229;$$

$$b_2 = \beta_2 \frac{\sigma_y}{\sigma_{x_2}} = -0{,}48045 \cdot \frac{1{,}75357}{28{,}6496} = -0{,}0294;$$

$$a = \bar{s} - b_1\bar{Y} - b_2\bar{W} = 3{,}8 - 0{,}1229 \cdot 40 + 0{,}0294 \cdot 47{,}4 = 0{,}2787.$$

Эти значения оценок совпадают с оценками, полученными ранее.

Показатели качества множественной регрессии. Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – коэффициента детерминации. Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, или оценивает тесноту совместного влияния факторов на результат. Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции:

$$R_{yx_1x_2\ldots x_p} = \sqrt{1 - \frac{\sigma^2{}_{\text{îñò}}}{\sigma^2{}_y}}\ .$$

Методика построения индекса множественной корреляции аналогична построению индекса корреляции для парной зависимости. Границы его изменения те же: от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов.

Если обратиться к линейному уравнению множественной регрессии в стандартизованном масштабе, то для расчета индекса множественной корреляции можно использовать формулу следующего вида:

$$R_{yx_1x_2\ldots x_p} = \sqrt{\sum \beta_{x_i} \cdot r_{yx_i}}$$

Формула индекса множественной корреляции для линейной регрессии получила название линейного коэффициента множественной корреляции, или совокупного коэффициента корреляции.

При линейной зависимости определение совокупного коэффициента корреляции возможно без построения регрессии и оценки её параметров, а с использованием только матрицы парных коэффициентов корреляции:

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

где Δr – определитель матрицы парных коэффициентов корреляции:

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_p} \\ r_{x_1y} & 1 & r_{x_1x_2} & \cdots & r_{x_1x_p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_py} & r_{x_px_1} & r_{x_px_2} & \cdots & 1 \end{vmatrix}$$

а Δr_{11} – определитель матрицы межфакторной корреляции:

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & \cdots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & r_{x_2x_3} & \cdots & r_{x_2x_p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_px_1} & r_{x_px_2} & r_{x_px_3} & \cdots & 1 \end{vmatrix}$$

Определитель матрицы межфакторной корреляции остаётся после вычеркивания из матрицы коэффициентов парной корреляции первого столбца и первой строки, что и соответствует матрице коэффициентов парной корреляции между факторами.

Проверка статистического качества оцененного уравнения регрессии проводится, с одной стороны, по статистической значимости параметров уравнения, а с другой стороны, по общему качеству уравнения регрессии. Кроме этого, проверяется выполнимость предпосылок МНК.

Сначала рассмотрим первые два вида проверок и связанные с ними вопросы. Некоторые предпосылки МНК и проверки их выполнимости будем рассматривать отдельно.

Для проверки общего качества уравнения регрессии используется коэффициент детерминации R^2, который в общем случае рассчитывается по формуле: $R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$. Он показывает, как и в парной регрессии, долю общей дисперсии y, объясненную уравнением регрессии. Его значения находятся между нулем и единицей. Чем ближе этот коэффициент к единице, тем больше уравнение регрессии объясняет поведение y.

Для множественной регрессии R^2 является неубывающей функцией числа объясняющих переменных. Добавление новой объясняющей переменной никогда не уменьшает значение R^2. Действительно, каждая следующая объясняющая переменная может лишь дополнить, но никак не сократить информацию, объясняющую поведение зависимой переменной.

В формуле расчета коэффициента детерминации используется остаточная дисперсия, которая имеет систематическую ошибку в сторону уменьшения, тем более значительную, чем больше параметров определяется в уравнении регрессии при заданном объёме наблюдений n. Если число параметров $(p+1)$ приближается к n, то остаточная дисперсия будет близка к нулю и коэффициент детерминации приблизится к единице даже при слабой связи факторов с результатом.

Поэтому в числителе и знаменателе делается поправка на число степеней свободы остаточной и общей дисперсии соответственно и рассчитывается скорректированный коэффициент детерминации:

$$\overline{R}^2 = 1 - \frac{\sum e_i^2 / (n-p-1)}{\sum (y_i - \overline{y})^2 / (n-1)}$$

Поскольку величина обычного коэффициента детерминации, как правило, увеличивается при добавлении объясняющей переменной к уравнению регрессии даже без достаточных на то оснований, скорректированный коэффициент детерминации компенсирует это увеличение путем наложения «штрафа» за увеличение числа независимых переменных. Перепишем формулу скорректированного коэффициента детерминации следующим образом:

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} = \frac{n-1}{n-p-1}R^2 - \frac{p}{n-p-1} = R^2 - \frac{p}{n-p-1}\left(1 - R^2\right)$$

По мере роста p увеличивается отношение $p/(n\text{-}p\text{-}1)$ и, следовательно, возрастает размер корректировки коэффициента R^2 в сторону уменьшения.

Очевидно, что $\overline{R}^2 < R^2$ при $p>1$. С ростом p $\overline{R}^2$ растет медленнее, чем R^2. Другими словами, он корректируется в сторону уменьшения с ростом числа объясняющих переменных. При этом $\overline{R}^2 = R^2$ только при $R^2=1$. $\overline{R}^2$ может даже принимать отрицательные значения (например, при $R^2=0$). Поэтому для корректировки формулы скорректированного коэффициента детерминации нет строгого математического обоснования.

Доказано, что $\overline{R}^2$ увеличивается при добавлении новой объясняющей переменной тогда и только тогда, когда t – статистика для этой переменной по модулю больше единицы. Из этого отнюдь не следует, как можно было бы предположить, что увеличение $\overline{R}^2$ означает улучшение спецификации уравнения. Тем не менее, добавление в модель новых факторов осуществляется до тех пор, пока растет скорректированный коэффициент детерминации.

Обычно приводятся данные как по R^2, так и по $\overline{R}^2$, являющиеся суммарными мерами общего качества уравнения регрессии. Однако не следует абсолютизировать значимость коэффициентов детерминации. Существует немало примеров неправильно построенных моделей, имеющих высокие коэффициенты детерминации. Поэтому коэффициент детерминации в настоящее время рассматривается лишь как один из ряда показателей, которые нужно проанализировать, чтобы уточнить строящуюся модель.

Анализ статистической значимости коэффициента детерминации проводится на основе проверки нуль – гипотезы H_0: $R^2=0$ против альтернативной гипотезы H_1: $R^2>0$. Для проверки данной гипотезы используется следующая F – статистика:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p}$$

Величина F при выполнении предпосылок МНК и при справедливости нуль – гипотезы имеет распределение Фишера. Из формулы расчета F-статистики видно, что показатели F и R^2 равны или не равны нулю одновременно. Если F=0, то R^2=0, и линия регрессии $y = \bar{y}$ является наилучшей по МНК, и, следовательно, величина y линейно не зависит от $x_1, x_2, ..., x_p$. Для проверки нуль – гипотезы при заданном уровне значимости α по таблицам критических точек распределения Фишера находится критическое значение $F_{табл}(\alpha;\ p;\ n\text{-}p\text{-}1)$. Если $F>F_{табл}$, нуль – гипотеза отклоняется, что равносильно статистической значимости R^2, т.е. $R^2>1$.

Эквивалентный анализ может быть предложен рассмотрением другой нуль – гипотезы, которая формулируется как $H_0 : \beta_1' = \beta_2' = ... = \beta_p' = 0$. Эту гипотезу можно назвать гипотезой об общей значимости уравнения регрессии. Если данная гипотеза не отклоняется, то делается вывод о том, что совокупное влияние всех p объясняющих переменных $x_1, x_2, ..., x_p$ на зависимую переменную y можно считать статистически несущественным, а общее качество уравнения регрессии невысоким.

Проверка такой гипотезы осуществляется на основе дисперсионного анализа сравнения объясненной и остаточной дисперсий, т.е. нуль – гипотеза формулируется как $H_0{:}D_{факт}=D_{ост}$ против альтернативной гипотезы $H_1{:}D_{факт}>D_{ост}$. При этом строится F – статистика:

$$F=\frac{\sum(\hat{y}_i-\bar{y})^2/p}{\sum(y_i-\hat{y}_i)^2/(n-p-1)}$$

Здесь в числителе – объясненная (факторная) дисперсия в расчете на одну степень свободы (число степеней свободы равно числу факторов, т.е. p). В знаменателе – остаточная дисперсия на одну степень свободы. Её число степеней свободы равно (n-p-1). Потеря (p+1) степени свободы связана с необходимостью решения системы (p+1) линейных уравнений при определении параметров эмпирического уравнения регрессии. Если учесть, что число степеней свободы общей дисперсии равно (n-1), то число степеней свободы объясненной дисперсии равна разности (n-1) – (n-p-1), т.е. p. Следует отметить, что выражение $F=\frac{\sum(\hat{y}_i-\bar{y})^2/p}{\sum(y_i-\hat{y}_i)^2/(n-p-1)}$ эквивалентно выражению $F=\frac{R^2}{1-R^2}\cdot\frac{n-p-1}{p}$. Это становится ясно, если числитель и знаменатель $F=\frac{\sum(\hat{y}_i-\bar{y})^2/p}{\sum(y_i-\hat{y}_i)^2/(n-p-1)}$ разделить на общую СКО:

$$F=\frac{\sum(\hat{y}_i-\bar{y})^2/\sum(y_i-\bar{y})^2}{\sum(y_i-\hat{y}_i)^2/\sum(y_i-\bar{y})^2}\cdot\frac{n-p-1}{p}=\frac{R^2}{1-R^2}\cdot\frac{n-p-1}{p}$$

Поэтому методика принятия или отклонения нуль – гипотезы для статистики $F=\frac{\sum(\hat{y}_i-\bar{y})^2/p}{\sum(y_i-\hat{y}_i)^2/(n-p-1)}$ ничем не отличается от таковой для статистики $F=\frac{R^2}{1-R^2}\cdot\frac{n-p-1}{p}$.

Анализ статистики F позволяет сделать вывод о том, что для принятия гипотезы об одновременном равенстве нулю всех коэффициентов линейной регрессии коэффициент детерминации R^2 должен существенно отличаться от нуля. Его критическое значение уменьшается при росте числа наблюдений и может стать сколь угодно малым.

Например, пусть при оценке регрессии с двумя объясняющими переменными по 30 наблюдениям R^2 =0,65. Тогда $F = \dfrac{0{,}65}{0{,}35} \cdot \dfrac{30-2-1}{2} \approx 25{,}07$

По таблицам критических точек распределения Фишера найдем F(0,05; 2; 27)=3,36; F(0,01; 2; 27)=5,49. Поскольку $F_{\text{набл}}$=25,05>$F_{\text{кр}}$ как при 5% - ном, так и при 1% - ном уровне значимости, то нулевая гипотеза в обоих случаях отклоняется. Если в той же ситуации R^2=0,4, то $F = \dfrac{0{,}65}{0{,}35} \cdot \dfrac{30-2-1}{2} \approx 25{,}07$. Предположение о незначимости связи отвергается и здесь.

Как и в случае парной регрессии, статистическая значимость параметров множественной линейной регрессии с p факторами проверяется на основе t – статистики: $t_{b_j} = \dfrac{b_j}{m_{b_j}}$ $\left(\text{или } t_a = \dfrac{a}{m_a} \right)$, где величина $m_{b_j}(m_a)$ называется стандартной ошибкой параметра $b_j(a)$. Она определяется так. Обозначим матрицу: $Z^{-1} = (X'X)^{-1}$, и в этой матрице обозначим j – й диагональный элемент как z_{jj}'. Тогда выборочная дисперсия эмпирического параметра регрессии равна: $m_{b_j}^2 = s^2 z_{jj}', \quad j = \overline{1,p}$, а для свободного члена выражение имеет вид: $m_a^2 = s^2 z_{00}'$, если считать, что в матрице Z^{-1} индексы изменяются от 0 до p. Здесь S^2 – несмещенная оценка дисперсии случайной ошибки ε: $s^2 = \dfrac{\sum e_i^2}{n-p-1}$.

Стандартные ошибки параметров регрессии равны: $m_{b_j} = \sqrt{m_{b_j}^2} \left(\text{или } m_a = \sqrt{m_a^2} \right)$.

Полученная по выражению $t_{b_j} = \dfrac{b_j}{m_{b_j}}$ $\left(\text{или } t_a = \dfrac{a}{m_a} \right) t$ – статистика для соответствующего параметра имеет распределение Стьюдента с числом степеней свободы (n-p-1). При требуемом уровне значимости α эта статистика сравнивается с критической точкой распределения Стьюдента $t(\alpha;\ n\text{-}p\text{-}1)$ (двухсторонней). Если $|t|>t(\alpha;\ n\text{-}p\text{-}1)$, то соответствующий параметр считается статистически значимым, и нуль – гипотеза в виде H_0:b_j=0 или H_0:a=0 отвергается. В противном случае ($|t|<t(\alpha;\ n\text{-}p\text{-}1)$) параметр считается статистически незначимым, и нуль – гипотеза не может быть отвергнута. Поскольку b_j не отличается значимо от нуля, фактор x_j линейно не связан с результатом. Его наличие среди объясняющих переменных не оправдано со статистической точки зрения. Не оказывая какого – либо серьёзного влияния на зависимую переменную, он лишь искажает реальную картину взаимосвязи. Поэтому после установления того факта, что коэффициент b_j статистически незначим, переменную x_j рекомендуется исключить из уравнения регрессии. Это не приведет к существенной потере качества модели, но сделает её более конкретной.

Строгую проверку значимости параметров можно заменить простым сравнительным анализом.

Если $|t| \le 1$, т.е. $b_j < m_{b_j}$, то коэффициент статистически незначим.

Если $1 < |t| \le 2$, т.е. $b_j < 2m_{b_j}$, то коэффициент относительно значим. В данном случае рекомендуется воспользоваться таблицей критических точек распределения Стьюдента.

Если $2 < |t| \le 3$, то коэффициент значим. Это утверждение является гарантированным при (n-p-1)>20 и $\alpha \ge 0{,}05$.

Если $|t|>3$, то коэффициент считается сильно значимым. Вероятность ошибки в данном случае при достаточном числе наблюдений не превосходит 0,001.

К анализу значимости коэффициента b_j можно подойти по – другому. Для этого строится интервальная оценка соответствующего коэффициента. Если задать уровень значимости α, то доверительный интервал, в который с вероятностью (1-α) попадает неизвестное значение параметра $\beta_j'(\alpha')$, определяется неравенством:

$$b_j - t(\alpha; n-p-1)\cdot m_{b_j} < \beta_j' < b_j + t(\alpha; n-p-1)\cdot m_{b_j}$$

или

$$a - t(\alpha; n-p-1)\cdot m_a < \alpha' < a + t(\alpha; n-p-1)\cdot m_a$$

Если доверительный интервал не содержит нулевого значения, то соответствующий параметр является статистически значимым, в противном случае гипотезу о нулевом значении параметра отвергать нельзя.

Мультиколлинеарность. Мультиколлинеарность - это линейная взаимосвязь двух или нескольких объясняющих переменных (x_1, x_2, … x_m). Если объясняющие переменные связаны строгой функциональной зависимостью, то говорят о совершенной мультиколлинеарности. Мультиколлинеарность не позволяет однозначно разделить вклады объясняющих переменных x1,x2,…xm в их влияние на зависимую переменную Y.

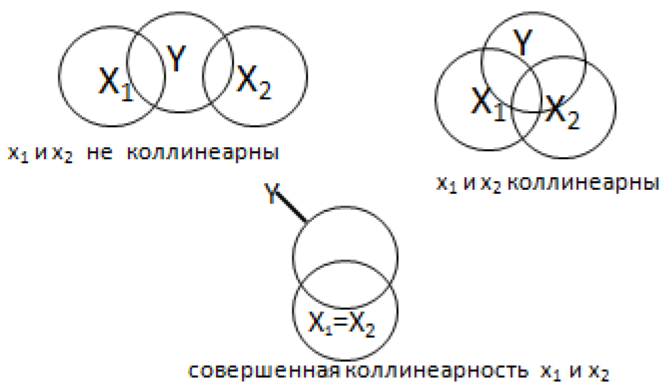


Рис.5.3. Диаграмма Венна

Коррелированность регрессоров обусловливает существенное усложнение процедуры поиска наилучшего уравнения регрессии, так как любое изменение состава регрессоров приводит к необходимости пересчитывать все параметры заново.Если факторы связаны строгой функциональной зависимостью, то это свидетельствует о полной (совершенной, строгой) мультиколлинеарности. Полная мультиколлинеарность не позволяет однозначно оценить параметры исходной модели и разделить вклады регрессоров в зависимую переменную Y. Наличие линейно связанных регрессоров относят к ошибкам спецификации. Такие ошибки при двух линейно связанных регрессорах встречаются крайне редко и легко могут быть выявлены при анализе матрицы парных коэффициентов корреляции.

Чаще возникают ошибки, обусловленные включением в модель факторов, один из которых является линейной комбинацией нескольких других. Так, при использовании количественных показателей, характеризующих часть какой-либо величины, в число объясняющих переменных нельзя включать все составляющие этой величины, так как при этом одну из них можно определить путем вычитания из этой величины значений остальных факторов. Например, в линейной регрессионной модели оборота банка (Y) недопустимым является одновременное использование в модели следующих независимых переменных: сумма кредитов, выданных юридическим лицам (X1), сумма кредитов, выданных физическим лицам (X2), общая сумма кредитов, выданных банком (X3=X1+X2). В регрессионной модели $\widehat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3$ увеличение значений коэффициентов при первых двух регрессорах на произвольную константу c и уменьшение на эту же константу значения коэффициента при третьем регрессоре не приведет к изменению значения зависимой переменной. Это означает, что при одних и тех же значениях регрессоров и зависимой переменной существует множество различных значений параметров уравнения.

Последствия мультиколлинеарности: увеличиваются стандартные ошибки оценок; уменьшаются t-статистики МНК-оценок регрессии; МНК-оценки чувствительны к изменениям данных; возможность неверного знака МНК-оценок; трудность в определении вклада независимых переменных в дисперсию зависимой переменной.

В реальных эконометрических исследованиях мультиколлинеарность чаще проявляется в стохастической форме, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь. Иногда такой вид мультиколлинеарности называют частичной (несовершенной, реальной, скрытой, неполной). Матрица $X'X$ в этом случае является неособенной (близкой к вырожденной), имеет полный ранг, но ее определитель очень мал, т.е. близок к нулю. Такие матрицы ещё называют плохо обусловленными.

Частичная мультиколлинеарность приводит к следующим последствиям:

Увеличение дисперсий оценок параметров. Это расширяет интервальные оценки и ухудшает их точность. Уменьшение t-статистик коэффициентов, что приводит к неоправданному выводу о значимости регрессоров. Неустойчивость МНК – оценок параметров и их дисперсий: небольшое изменение исходных данных (добавление или исключение одного – двух наблюдений) будет приводить к значительному изменению этих оценок. Возможность получения неверного (с точки зрения теории) знака у параметра регрессии или неоправданно большого значения этого параметра. В результате получаются значительные средние квадраты отклонения коэффициентов регрессии a, b1, b2, b3 … bp и оценка их значимости по t-критерию Стьюдента не имеет смысла, хотя в целом регрессионная модель может оказаться значимой по F-критерию.

Наиболее простой формой сильной взаимосвязи факторов является высокая парная корреляция регрессоров. Она может быть выявлена при анализе матрицы парных коэффициентов корреляции. Обычно факторы считаются тесно связанными, если значения выборочных парных коэффициентов корреляции $\left|r_{x_i x_j}\right| > (0{,}7 \ldots 0{,}8)$. При наличии такой тесной связи для какой-либо пары признаков обычно рекомендуется не включать в модель один из них, если это допустимо с точки зрения корректности модели.

Действительная мультиколлинеарность в полном смысле слова возникает при наличии тесной взаимосвязи множества независимых переменных. Она может и не обнаруживаться по матрице парных коэффициентов корреляции. В отсутствие тесной корреляционной связи одного из признаков с каждым из остальных может наблюдаться тесная связь с их совокупностью. Такую связь можно выявить путем углубленного корреляционного анализа. Он состоит в том, что при значениях множественного коэффициента корреляции какого-либо j – го независимого фактора с остальными регрессорами модели $R_j \geq (0{,}7 \ldots 0{,}8)$ можно говорить о наличии проблемы мультиколлинеарности. Основная проблема заключается в том, что расчет множественных коэффициентов корреляции каждого из регрессоров с совокупностью остальных факторов модели может не дать нужного результата, поскольку наличие мультиколлинеарности в этой совокупности искажает и результат оценки степени взаимосвязи независимых переменных.

Поскольку заранее корреляционная структура данных, как правило, неизвестна, это приводит к необходимости рассчитывать большое число множественных коэффициентов корреляции, начиная с анализа взаимосвязи одного признака со всеми возможными парами из остальных, затем с тройками признаков и т.д. Такой анализ становится очень трудоемким и редко используется на практике.

Признаки мультиколлинеарности: высокий R^2; близкая к 1 парная корреляция между малозначимыми независимыми переменными; высокие частные коэффициенты корреляции; сильная дополнительная регрессия между независимыми переменными.

Методы устранения мультиколлинеарности: исключение из модели коррелированных переменных (при отборе факторов); сбор дополнительных данных или новая выборка; изменение спецификации модели; использование предварительной информации о параметрах; преобразование переменных.

Мультиколлинеарность чаще всего обнаруживает себя в ходе регрессионного анализа. К ее признакам можно отнести следующие:

1) значительные изменения коэффициентов при регрессорах при изменениях состава регрессоров и объектов, входящих в выборку;

2) незначимость большинства или всех коэффициентов при значимости уравнения в целом;

3) чрезмерно высокие или противоречащие экономической теории значения коэффициентов регрессионной модели.

Таким образом, точных количественных критериев для определения наличия или отсутствия мультиколлинеарности не существует. Тем не менее, ее наличие можно обнаружить с помощью:

1. Анализа корреляционной матрицы между объясняющими переменными и выявлении пар переменных, имеющих высокие коэффициенты корреляции.

2. Расчета множественных коэффициентов корреляции (коэффициентов детерминации) между одной из объясняющих переменных и некоторой группы из них. Наличие высокого множественного коэффициента детерминации свидетельствует о мультиколлинеарности.

3. Проверки чувствительности (устойчивости) оценок коэффициентов к небольшим изменениям исходных данных.

4. Исследования матрицы $(X'X)$. Если определитель матрицы $(X'X)$ либо ее минимальное собственное значение λ_{min} близки к нулю, то это говорит о наличии мультиколлинеарности. Об этом же может свидетельствовать и значительное отклонение максимального собственного значения $\lambda_{\max}$ матрицы $(X'X)$ от ее минимального собственного значения λ_{min}.

Одним из способов устранения мультиколлинеарности является исключение переменных из модели. Самым простым, но далеко не всегда возможным является способ, когда из двух объясняющих переменных, имеющих высокий коэффициент корреляции (обычно больше 0,8), одну переменную исключают из рассмотрения. При этом в первую очередь на основании экономических соображений решают, какую переменную оставить, а какую удалить из анализа.

Если с экономической точки зрения ни одной из переменных нельзя отдать предпочтение, то оставляют ту из двух переменных, которая имеет больший коэффициент корреляции с зависимой переменной.

Более углубленный анализ регрессоров можно получить, используя метод дополнительной регрессии. Его суть заключается в том, что для выявления списка зависимых регрессоров проводится дополнительная регрессия – регрессия каждого независимого фактора X_j, $j=1,2,\dots p$ на оставшиеся независимые факторы. Стандартным способом, на основе F-статистики, проверяется статистическая значимость коэффициентов детерминации R_j^2 дополнительных регрессий:

$$F_j = \frac{R_j^2}{1-R_j^2} \cdot \frac{n-p}{p-1}$$

где n – число наблюдений, p – число независимых переменных в первоначальной спецификации регрессионной модели. Статистика F_j имеет распределение Фишера с параметрами: $\nu_1 = p-1,\ \nu_2 = n-p$. Если коэффициент R_j^2 статистически не значим, то регрессор X_j не приводит к мультиколлинеарности и его оставляют в списке переменных модели. В противном случае рекомендуется исключить его из списка.

В ряде случаев можно попытаться изменить *спецификацию модели*: либо изменить форму модели, либо добавить объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на зависимую переменную. В результате уменьшается сумма квадратов отклонений, а, следовательно, сокращается стандартная ошибка регрессии. В свою очередь это приводит к уменьшению стандартных ошибок параметров модели.

Вопросы и задания для самоконтроля

1. Как записывается эмпирическое уравнение линейной модели множественной регрессии?

2. Что измеряют коэффициенты регрессии линейной модели множественной регрессии?

3. Какие этапывключает алгоритм определения коэффициентов множественной линейной регрессии по МНК в матричной форме?

4. Какие требования предъявляются к факторам для их включения их в модель множественной регрессии?

5. Как интерпретируются коэффициенты регрессии линейной модели потребления?

6. Какой смысл приобретает сумма коэффициентов регрессии в производственных фукнциях?

7. Как в линейной модели множественной регрессии, записанной в стандартизованном виде, сравнить факторы по силе их воздействия на результат?

8. Как связаны стандартизованные коэффициенты регрессии с натуральными?

9. Как корректируется коэффициент детерминации?

10. Каково назначение частной корреляции при построении модели множественной регрессии?

11. Как проверяется адекватность линейной модели множественной регрессии в целом?

Задача 1. Получены следующие величины:

$\bar{y} = 15{,}0$; $\bar{x}_1 = 6{,}5$; $\bar{x}_2 = 12{,}0$; $\sigma_y = 4{,}0$; $\sigma_{x_1} = 2{,}5$; $\sigma_{x_2} = 3{,}5$; $r_{yx_1} = 0{,}63$; $r_{yx_2} = 0{,}78$; $r_{x_1x_2} = 0{,}52$.

Задание: записать регрессию y на x_1 и x_2 в стандартизованной и естественной формах.

Задача 2. Уравнение регрессии, построенное по 15 наблюдениям, имеет вид:

$$\tilde{y} = 12{,}4 - 9{,}6x_1 + ?\,x_2 - 6{,}3x_3$$

m_b (?) (3,2) (0,12) (?)

t_b (1,55) (?) (4,0) (–3,15).

Задание: определить пропущенные значения и построить доверительный интервал для β_3 с вероятностью 0,99.

Задача 3. Уравнение регрессии в стандартизованной форме имеет вид

$$t_y = 0{,}37t_{x_1} - 0{,}52t_{x_2} + 0{,}43t_{x_3} .$$

При этом коэффициенты вариации равны:

$$V_y = 18\%,\ V_{x_1} = 25\%,\ V_{x_2} = 38\%,\ V_{x_3} = 30\%.$$

Задание: определить частные коэффициенты эластичности.

Лекция 2(2)

Тема 8, 9. Гетероскедастичность и автокорреляция в остатках регрессии

Вопросы для изучения

1. Понятие и последствия гетероскедастичности.
2. Обнаружение и устранение гетероскедастичности.
3. Понятие и последствия автокорреляции.
4. Обнаружение и устранение автокорреляции.

Аннотация. Данная тема раскрывает способы проверки соблюдения второй и третьей предпосылок МНК в остатках регрессии.

Ключевые слова. Гетероскедастичность, автокорреляция, остатки регрессии.

Методические рекомендации по изучению темы

• Тема содержит лекционную часть, где даются общие положения по теме;

• В качестве самостоятельной работы предлагается ознакомиться с решениями типовых задач, презентацией и ответить на вопросы для изучения.

• Для проверки усвоения темы имеется тест для самоконтроля.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11766

2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none) С. 92-106.

3. Валентинов, В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 202-229.

5.Эконометрика. Практикум: [Электронный ресурс] Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов.знание, 2014. - 329 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 197-244.

Понятие и последствия гетероскедастичности. Гетероскедастичностью остатков называется нарушение 2 предпосылки МНК о постоянстве дисперсий случайных отклонений. Если предпосылка МНК о том, что $D(\varepsilon_i)=D(\varepsilon_j)=\sigma^2$ соблюдена, то имеет место гомоскедастичность случайных отклонений. Последствия гетероскедастичности: МНК-оценки сохраняют свойства несмещенности и линейности, но теряют свойство эффективности; дисперсии МНК-оценок смещены; t-статистика и F-статистика завышены. В качестве примера реальной гетероскедастичности можно привести то, что люди с большим доходом не только тратят в среднем больше, чем люди с меньшим доходом, но и разброс в их потреблении также больше, поскольку они имеют больше простора для распределения дохода.

В ряде случаев, зная характер исходных данных, можно предвидеть гетероскедастичность и попытаться устранить её ещё на стадии спецификации. Однако значительно чаще эту проблему приходится решать после построения уравнения регрессии.

Обнаружение и устранение гетероскедастичности. Графическое построение отклонений от эмпирического уравнения регрессии позволяет визуально определить наличие гетероскедастичности. В этом случае по оси абсцисс откладываются значения объясняющей переменной x_i (для парной регрессии) либо линейную комбинацию объясняющих переменных:

$$\hat{y}_i = a + b_1 x_{i1} + \ldots + b_p x_{ip}, \quad i = \overline{1, n}$$

(для множественной регрессии), а по оси ординат либо отклонения e_i, либо их квадраты $e_i^2, \quad i = \overline{1, n}$.

Если все отклонения e_i^2 находятся внутри горизонтальной полосы постоянной ширины, это говорит о независимости дисперсий e_i^2 от значений объясняющей переменной и выполнимости условия гомоскедастичности.

В других случаях наблюдаются систематические изменения в соотношениях между значениями $\hat{y}_i$ и квадратами отклонений e_i^2. Такие ситуации отражают большую вероятность наличия гетероскедастичности для рассматриваемых статистических данных. В настоящее время для определения гетероскедастичности разработаны специальные тесты и критерии для них.

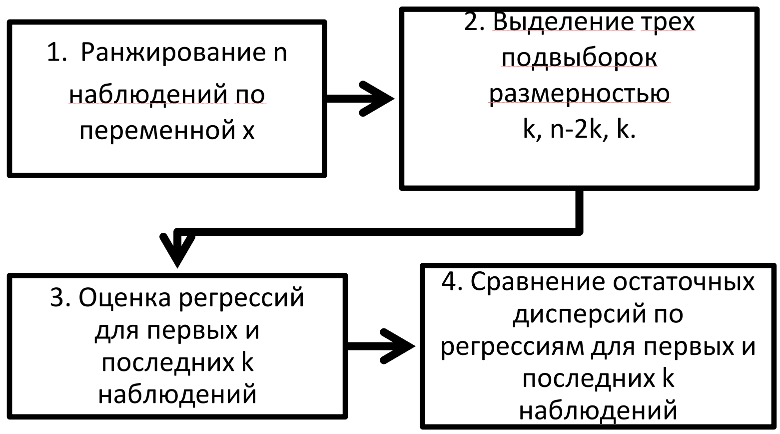


Рис. 8.1. Тест Голдфелда-Квандта

F-статистика для сравнения дисперсий:

$$S^2{}_1 = \sum_{i=1}^{k} e_i^{\,2}; S^2{}_3 = \sum_{i=n-k+1}^{n} e_i^{\,2},$$
$$H_0 : S^2{}_3 = S^2{}_1 (\textit{гомоскедастичность})$$
$$H_1 : S^2{}_3 > S^2{}_1 (\textit{гетероскедастичность})$$
$$F = \frac{S^2{}_3 /(k-m-1)}{S^2{}_1 /(k-m-1)},$$
$$F > F_{\alpha,m,k-m-1} \Rightarrow H_1$$

Тест ранговой корреляции Спирмена. При использовании данного теста предполагается, что дисперсия отклонений будет либо увеличиваться, либо уменьшаться с увеличением значений *x*. Поэтому для регрессии, построенной по МНК, абсолютные величины отклонений $|e_i|$ и значения x_i будут коррелированы. t- статистика для проверки значимости $r_{x,e}$:

$$r_{x,e} = 1 - 6 \cdot (\sum d_i^{\,2} / n(n^2 - 1))$$
$$H_0 : r_{x,e} = 0 (\textit{гомоскедастичность})$$
$$H_1 : r_{x,e} \neq 0 (\textit{гетероскедастичность})$$
$$t = \frac{r_{x,e} \cdot \sqrt{n-2}}{\sqrt{1 - r^2{}_{x,e}}},$$
$$t > t_{\alpha,n-2} \Rightarrow H_1$$

Для устранения гетероскедастичности в случае, если дисперсии отклонений известны для каждого наблюдения, применяется метод взвешенных наименьших квадратов (ВНК). Гетероскедастичность устраняется, если разделить каждое наблюдаемое значение на соответствующее ему значение дисперсии:

$$y = \alpha + \beta \cdot x + \varepsilon$$
$$\frac{y}{\sigma} = \alpha \cdot \frac{1}{\sigma} + \beta \cdot \frac{x}{\sigma} + \frac{\varepsilon}{\sigma}$$
$$y^* = \alpha \cdot z + \beta \cdot x^* + v$$

Если дисперсии отклонений неизвестны для каждого наблюдения, то предполагается, что дисперсии σ^2_e пропорциональны x_i

$$\sigma^2{}_i = \sigma^2 x_i$$

$$\frac{y_i}{\sqrt{x_i}} = \alpha \cdot \frac{1}{\sqrt{x_i}} + \beta \cdot \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}}$$

$$\frac{y_i}{\sqrt{x_i}} = \alpha \cdot \frac{1}{\sqrt{x_i}} + \beta \cdot \sqrt{x_i} + v_i$$

$$y* = \alpha \cdot z + \beta \cdot x* + v_i$$

Дисперсии σ^2_e пропорциональны x^2_i

$$\sigma^2{}_i = \sigma^2 \cdot x^2{}_i$$

$$\frac{y_i}{x_i} = \alpha \cdot \frac{1}{x_i} + \beta \cdot \frac{x_i}{x_i} + \frac{\varepsilon_i}{x_i}$$

$$\frac{y_i}{x_i} = \alpha \cdot \frac{1}{x_i} + \beta + v_i$$

$$y* = \alpha \cdot z + \beta + v_i$$

Таким образом, наблюдения с наименьшими дисперсиями получают наибольшие «веса», а наблюдения с наибольшими дисперсиями – наименьшие «веса». Поэтому наблюдения с меньшими дисперсиями отклонений будут более значимыми при оценке параметров регрессии, чем наблюдения с большими дисперсиями. При этом повышается вероятность получения более точных оценок. В этом заключается смысл ВНК.

Полученные по МНК оценки параметров модели можно использовать в первоначальной модели.

Для применения ВНК необходимо знать фактические значения дисперсий отклонений σ_i^2. На практике такие значения известны крайне редко. Поэтому, чтобы применить ВНК, необходимо сделать реалистические предположения о значениях σ_i^2. Чаще всего предполагается, что дисперсии отклонений пропорциональны или значениям x_i, или значениям x_i^2.

Если в уравнении регрессии присутствует несколько объясняющих переменных, вместо конкретной переменной x_j используется исходное уравнение множественной регрессии

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p \ ,$$

т.е. фактически линейная комбинация факторов. В этом случае получают следующую регрессию: $\frac{y_i}{\sqrt{\hat{y}_i}} = a\frac{1}{\sqrt{\hat{y}_i}} + b_1\frac{x_{i1}}{\sqrt{\hat{y}_i}} + \ldots + b_p\frac{x_{ip}}{\sqrt{\hat{y}_i}} + \frac{\varepsilon_i}{\sqrt{\hat{y}_i}}$.

Понятие и последствия автокорреляции. Автокорреляцией остатков называется нарушение третьей предпосылки МНК о независимости случайного отклонения ε i от отклонений во всех других наблюдениях. Если предпосылка МНК о том, что cov(εi,εj)=0, соблюдена, то автокорреляция случайных отклонений отсутствует. Чаще всего положительная автокорреляция вызывается направленным постоянным воздействием некоторых не учтенных в регрессии факторов. Например, при исследовании спроса *y* на прохладительные напитки в зависимости от дохода *x* на трендовую зависимость накладываются изменения спроса в летние и зимние периоды. Аналогичная картина может иметь место в макроэкономическом анализе с учетом циклов деловой активности. Автокорреляция остатков обычно встречается при использовании данных временных рядов. В перекрестных данных наличие автокорреляции бывает редко. Положительная автокорреляция имеет место, когда rεi,εj>0. Отрицательная автокорреляция имеет место, когда rεi,εj<0. Отрицательная автокорреляция фактически означает, что за положительным отклонением следует отрицательное и наоборот. Такая ситуация может иметь место, если ту же зависимость между спросом на прохладительные напитки и доходами рассматривать не ежемесячно, а раз в сезон (зима–лето).

Последствия автокорреляции: МНК-оценки сохраняют свойства несмещенности и линейности, но теряют свойство эффективности; дисперсии МНК-оценок смещены в сторону занижения; t-статистика и F-статистика завышены.

Обнаружение и устранение автокорреляции. Методы обнаружения автокорреляции: графический анализ остатков; критерий Дарбина-Уотсона; метод рядов.

Метод рядов. По этому методу последовательно определяются знаки отклонений $e_t, t = \overline{1,n}$ от регрессионной зависимости. Например, имеем при 20 наблюдениях

$$(-----)(+++++++)(---)(++++)(-)\ .$$

Ряд определяется как непрерывная последовательность одинаковых знаков. Количество знаков в ряду называется длиной ряда. Если рядов слишком мало по сравнению с количеством наблюдений *n*, то вполне вероятна положительная автокорреляция. Если же рядов слишком много, то вероятна отрицательная автокорреляция.

Пусть n – объём выборки, n_1 – общее количество положительных отклонений; n_2 – общее количество отрицательных отклонений; k – количество рядов. В приведенном примере n=20, n_1=11, n_2=5.

При достаточно большом количестве наблюдений (n_1>10, n_2>10) и отсутствии автокорреляции СВ k имеет асимптотически нормальное распределение, в котором

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

$$D(k) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

Тогда, если

$$M(k) - u_{\alpha/2} \cdot \sqrt{D(k)} < k < M(k) + u_{\alpha/2} \cdot \sqrt{D(k)},$$

то гипотеза об отсутствии автокорреляции не отклоняется. Если $k \le M(k) - u_{\alpha/2} \cdot \sqrt{D(k)}$, то констатируется положительная автокорреляция; в случае $k \ge M(k) + u_{\alpha/2} \cdot \sqrt{D(k)}$ признается наличие отрицательной автокорреляции.

Для небольшого числа наблюдений (n_1<20, n_2<20) были разработаны таблицы критических значений количества рядов при n наблюдениях. В одной таблице в зависимости от n_1 и n_2 определяется нижняя граница k_1 количества рядов, в другой – верхняя граница k_2. Если $k_1<k<k_2$, то говорят об отсутствии автокорреляции. Если $k \le k_1$, то говорят о положительной автокорреляции. Если $k \ge k_2$,то говорят об отрицательной автокорреляции. Например, для приведенных выше данных k_1=6, k_2=16 при уровне значимости 0,05. Поскольку $k=5<k_1=6$, определяем положительную автокорреляцию.

Критерий Дарбина-Уотсона:

$$DW = \frac{\sum_{n=2}^{N}(e_i - e_{i-1})^2}{\sum_{n=1}^{N} e_i^2}$$

$$DW \approx 2 \cdot (1 - r_{e_i, e_{i-1}}); 0 \le DW \le 4$$

$$r_{e_i, e_{i-1}} \approx 0 \Rightarrow DW \approx 2$$

$$r_{e_i, e_{i-1}} \approx 1 \Rightarrow DW \approx 0(\text{"+"}\ \textit{автокорреляция})$$

$$r_{e_i, e_{i-1}} \approx -1 \Rightarrow DW \approx 4(\text{"–"}\ \textit{автокорреляция})$$

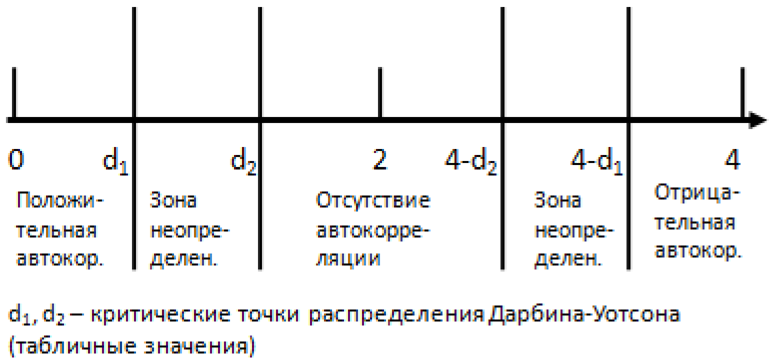


Рис. 8.2. Проверка гипотезы об автокорреляции остатков по DW-критерию

Можно показать, что статистика *DW* тесно связана с коэффициентом автокорреляции первого порядка: $r_{e_{t-1}e_t} = \dfrac{\sum_{t=2}^{n} e_{t-1}e_t}{\sqrt{\sum_{t=1}^{n-1} e_t^2 \sum_{t=2}^{n} e_{t-1}^2}}$.

Связь выражается формулой: $DW \approx 2\left(1 - r_{e_{t-1}e_t}\right)$.

Отсюда вытекает смысл статистического анализа автокорреляции. Поскольку значения r изменяются от –1 до +1, DW изменяется от 0 до 4. Когда автокорреляция отсутствует, коэффициент автокорреляции равен нулю, и статистика DW равна 2. DW=0 соответствует положительной автокорреляции, когда выражение в скобках равно нулю $(r = +1)$. При отрицательной автокорреляции $(r = -1)$. DW=4, и выражение в скобках равно двум.

Ограничения критерия Дарбина-Уотсона:

1. Критерий DW применяется лишь для тех моделей, которые содержат свободный член.

2. Предполагается, что случайные отклонения определяются по итерационной схеме $e_t = \rho e_{t-1} + v_t$, называемой авторегрессионной схемой первого порядка $AR(1)$. Здесь v_t – случайный член.

3. Статистические данные должны иметь одинаковую периодичность (не должно быть пропусков в наблюдениях).

4. Критерий Дарбина – Уотсона не применим к авторегрессионным моделям вида: $y_t = a + b_1 x_{t1} + \ldots + b_p x_{tp} + c y_{t-1} + e_t$, которые содержат в числе факторов также зависимую переменную с временным лагом (запаздыванием) в один период.

Автокорреляция чаще всего вызывается неправильной спецификацией модели. Поэтому следует попытаться скорректировать саму модель, в частности, ввести какой – нибудь неучтенный фактор или изменить форму модели (например, с линейной на полулогарифмическую или гиперболическую). Если все эти способы не помогают и автокорреляция вызвана какими – то внутренними свойствами ряда {et}, можно воспользоваться преобразованием, которое называется авторегрессионной схемой первого порядка AR(1).

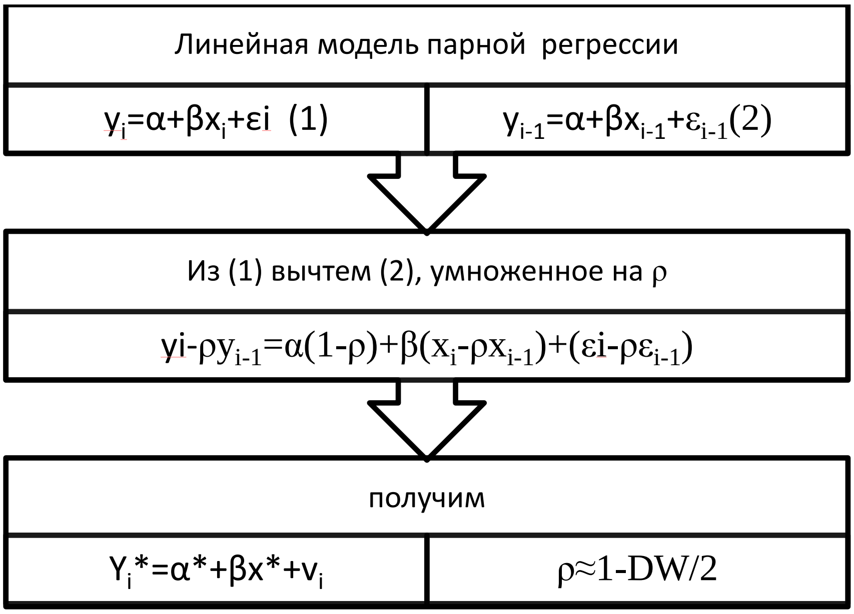


Рис.8.3. Авторегрессионное преобразование

Поскольку случайные отклонения v_t удовлетворяют предпосылкам МНК, оценки a^* и b будут обладать свойствами наилучших линейных несмещенных оценок. По преобразованным значениям всех переменных с помощью обычного МНК вычисляются оценки параметров a^* и b, которые затем можно использовать в регрессии (71).

Однако способ вычисления преобразованных переменных (75) приводит к потере первого наблюдения, если нет информации о предшествующих наблюдениях. Это уменьшает на единицу число степеней свободы, что при больших выборках не очень существенно, однако при малых выборках приводит к потере эффективности. Тогда первое наблюдение восстанавливается с помощью поправки Прайса – Уинстена:

$$x_1^* = \sqrt{1-\rho^2} \cdot x_1,$$
$$y_1^* = \sqrt{1-\rho^2} \cdot y_1$$

Авторегрессионное преобразование может быть обобщено на произвольное число объясняющих переменных, т.е. использовано для уравнения множественной регрессии.

Вопросы и задания для самоконтроля

1. В чем суть гомоскедастичности и гетероскедастичности? Каковы последствия гетероскедастичности?
2. Действительно ли, вследствие гетероскедастичности оценки перестают быть эффективными и состоятельными?
3. Какие критерии могут быть использованы для проверки гипотезы о гомоскедастичности регрессионных остатков?
4. В чем заключается тест Спирмена?
5. Какова схема теста Голдфелда-Квандта?
6. Каково предположение теста Парка?
7. Что такое автокорреляция случайных отклонений?
8. Каковы основные причины и последствия автокорреляции?
9. Что такое автокорреляционная функция?
10. Какова основная идея метода рядов при обнаружении автокорреляции?
11. Как проводится тест Дарбина-Уотсона?
12. В чем состоит авторегрессионная схема 1-го порядка?

Задача 1. Заданы следующие значения остатков линейной модели:

Ранг x_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
e_i	-1	2	-3	2	0	-3	3	1	-2	-4	5	-11	8	-20	12	-21	18	14

Задание: установить, имеется ли гетероскедастичность по тесту ранговой корреляции Спирмена на уровне значимости $\alpha = 0{,}05$.

Задача 2. По статистическим данным за 20 лет построено уравнение регрессии между ценой бензина и объемом продаж бензина, $d = DW = 0{,}71$.

Задание: ответить на вопросы: будет ли иметь место автокорреляция остатков? Что могло послужить причиной автокорреляции?

Лекция 3(1)

Тема 15. Модели одномерных временных рядов

Вопросы для изучения:

1. Понятие временного ряда и его основные компоненты.
2. Построение аддитивной модели.
3. Построение мультипликативной модели.

Аннотация. Данная тема раскрывает порядок построения аддитивных и мультипликативных моделей одномерных временных рядов.

Ключевые слова. Тренд, сезонные и случайные колебания, аддитивная модель, мультипликативная модель.

Методические рекомендации по изучению темы

•Тема содержит лекционную часть, где даются общие представления по теме.

•В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы.

•В качестве самостоятельной работы предлагается выполнить практические задания.

•Для проверки усвоения темы имеется тест для самоконтроля.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11766

2. Валентинов В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 242-261.

3. Эконометрика: учебник / И. И. Елисеева. – М.: Проспект, 2010. – 288 с. С.128-183.

4. Электронныйкурс “Time Series Econometrics”, Princeton University, URL:
http://sims.princeton.edu/yftp/Times05/;https://blackboard.princeton.edu/webapps/portal/frameset.jsp?tab_group=courses&url=%2Fwebapps%2Fblackboard%2Fexecute%2FcourseMain%3Fcourse_id%3D_52968_1.

Понятие временного ряда и его основные компоненты. Временнóй ряд – это совокупность значений какого – либо показателя за несколько последовательных моментов или периодов времени. Каждое значение (уровень) временного ряда формируется под воздействием большого числа факторов, которые можно условно разделить на три группы: факторы, формирующие тенденцию ряда; факторы, формирующие циклические колебания ряда; случайные факторы.

Тенденция характеризует долговременное воздействие факторов на динамику показателя. Тенденция может быть возрастающей или убывающей.

Циклические колебания могут носить сезонный характер или отражать динамику конъюнктуры рынка, а также фазу бизнес – цикла.

Реальные данные часто содержат все три компоненты. В большинстве случаев временной ряд можно представить как сумму или произведение трендовой (T), циклической (S) и случайной (E) компонент. В случае суммы имеет место аддитивная модель временного ряда:

$$y = T + S + E, \quad (1)$$

в случае произведения – мультипликативная модель:

$$y = T \cdot S \cdot E. \quad (2)$$

Основная задача эконометрического исследования отдельного временного ряда – получение количественного выражения каждой из компонент и использование этой информации для прогноза будущих значений ряда или построение модели взаимосвязи двух или более временных рядов.

Сначала рассмотрим основные подходы к анализу отдельного временного ряда. Такой ряд может содержать, помимо случайной составляющей, либо только тенденцию, либо только сезонную (циклическую) компоненту, либо все компоненты вместе. Для того, чтобы выявить наличие той или иной неслучайной компоненты, исследуется корреляционная зависимость между последовательными уровнями временного ряда, или автокорреляция уровней ряда. Основная идея такого анализа заключается в том, что при наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих.

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

Коэффициент автокорреляции уровней ряда первого порядка измеряет зависимость между соседними уровнями ряда t и $t-1,$ т.е. при лаге 1.

Он вычисляется по следующей формуле:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}}, \tag{3}$$

где в качестве средних величин берутся значения:

$$\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1}; \qquad \bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1}. \tag{4}$$

В первом случае усредняются значения ряда, начиная со второго до последнего, во втором случае - значения ряда с первого до предпоследнего.

Формулу (3) можно представить как формулу выборочного коэффициента корреляции:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \tag{5}$$

где в качестве переменной x берется ряд $y_2, y_3, \dots, y_n$, а в качестве переменной y – ряд $y_1, y_2, \dots, y_{n-1}$.

Если значение коэффициента (3) близко к единице, это указывает на очень тесную зависимость между соседними уровнями временного ряда и о наличии во временном ряде сильной линейной тенденции.

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

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}}, \qquad (6)$$

где в качестве одной средней величины берут среднюю уровней ряда с третьего до последнего, а в качестве другой - среднюю всех уровней ряда, кроме последних двух:

$$\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}; \quad \bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}. \qquad (7)$$

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют лагом. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Для обеспечения статистической достоверности максимальный лаг, как считают некоторые известные эконометристы, не должен превышать четверти общего объема выборки.

Коэффициент автокорреляции строится по аналогии с линейным коэффициентом корреляции, и поэтому он характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. По нему можно судить о наличии линейной или близкой к линейной тенденции. Однако для некоторых временных рядов с сильной нелинейной тенденцией (например, параболической или экспоненциальной), коэффициент автокорреляции уровней ряда может приближаться к нулю.

Кроме того, по знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных имеют положительную автокорреляцию уровней, однако при этом не исключается убывающая тенденция.

Последовательность коэффициентов автокорреляции уровней различных порядков, начиная с первого, называется автокорреляционной функцией временного ряда. График зависимости ее значений от величины лага называется коррелограммой. Анализ автокорреляционной функции и коррелограммы помогает выявить структуру ряда. Здесь уместно привести следующие качественные рассуждения.

Если наиболее высоким является коэффициент автокорреляции первого порядка, очевидно, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, ряд содержит циклические колебания с периодичностью в τ моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, то либо ряд не содержит тенденции и циклических колебаний и имеет только случайную составляющую, либо ряд содержит сильную нелинейную тенденцию, для исследования которой нужно провести дополнительный анализ.

В случае, если при анализе структуры временного ряда обнаружена только тенденция и отсутствуют циклические колебания (случайная составляющая присутствует всегда), следует приступать к моделированию тенденции. Если же во временном ряде имеют место и циклические колебания, прежде всего следует исключить именно циклическую составляющую, и лишь затем приступать к моделированию тенденции. Выявление тенденции состоит в построении аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называют аналитическим выравниванием временного ряда.

Зависимость от времени может принимать разные формы, поэтому для её формализации используют различные виды функций:

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

гипербола: $\hat{y}_t = a + b / t$;

экспоненциальный тренд: $\hat{y}_t = e^{a+b\cdot t}$ (или $\hat{y}_t = a \cdot b^t$);

степенной тренд: $\hat{y}_t = a \cdot t^b$;

параболический тренд второго и более высоких порядков:

$\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + ... + b_k \cdot t^k$.

Параметры каждого из трендов можно определить обычным МНК, используя в качестве независимой переменной время $t = 1, 2, ..., n$, а в качестве зависимой переменной – фактические уровни временного ряда y_t (или уровни за вычетом циклической составляющей, если таковая была обнаружена). Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Существует несколько способов определения типа тенденции. Чаще всего используют качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени, расчет некоторых основных показателей динамики. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни yt и yt-1 тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, если ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации $\overline{R}^2$ и выбора уравнения тренда с максимальным значением этого коэффициента. Реализация этого метода относительно проста при компьютерной обработке данных.

При анализе временных рядов, содержащих сезонные или циклические колебания, наиболее простым подходом является расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временно́го ряда в форме (1) или (2).

Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель (1), в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель (2), которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной модели.

1 шаг. *Выравнивание уровней ряда.* Просуммируем уровни ряда за каждые четыре квартала со сдвигом на один момент времени. Разделив полученные суммы на 4, найдем скользящие средние. Найдем центрированные скользящие средние как средние значения из двух последовательных скользящих средних.

2 шаг. *Расчет сезонной компоненты S.* Найдем разность между уровнями и центрированными скользящими средними. Расчет средней оценки сезонной компоненты для каждого квартала за все годы. Расчет скорректированной сезонной компоненты.Моделирование сезонных колебаний:Аддитивная модель: $Y_t = T_t + S_t + e_t$.

Оценка сезонной компоненты за каждый квартал: $s_t = y_t - \bar{y}_t$. Средняя оценка сезонной компоненты для квартала за все годы: $\bar{S}_t = \frac{\sum s_t}{n}$. Скорректированная сезонная компонента: $S_t = \bar{S}_t - k; k = \frac{\sum_{t=1}^{4} \bar{S}_t}{4}$

3 шаг. *Устранение сезонной компоненты S*.Вычтем скорректированное значение сезонной компоненты из каждого уровня исходного временного ряда. Получим: T+E=Y-S.

4 шаг. *Расчет значений тренда*. Проведем аналитическое выравнивание ряда (T+E) с помощью линейного тренда. Рассчитаем значения T для каждого момента времени по уравнению тренда.

5 шаг. *Расчет значений T+S*. Прибавим к уровням T значения сезонной компоненты (S) для соответствующих кварталов.

6 шаг. *Расчет абсолютной ошибки*. Выполним расчет ошибки для каждого уровня ряда по формуле: E=Y-(T+S). Расчет суммы квадратов абсолютных ошибок и ее сравнение с общей суммой квадратов отклонений уровней ряда.

Построение мультипликативной модели.

1 шаг. *Выравнивание уровней ряда*. Просуммируем уровни ряда за каждые четыре квартала со сдвигом на один момент времени. Разделив полученные суммы на 4, найдем скользящие средние. Найдем центрированные скользящие средние как средние значения из двух последовательных скользящих средних.

2 шаг. *Расчет сезонной компоненты S*. Найдем оценки сезонной компоненты как частное от деления уровней на центрированные скользящие средние. Расчет средней оценки сезонной компоненты для каждого квартала за все годы. Расчет скорректированной сезонной компоненты. Моделирование сезонных колебаний: Мультипликативная модель: $Y_t = T_t \cdot S_t \cdot e_t$.

Оценка сезонной компоненты за каждый квартал: $s_t = \frac{y_t}{\bar{y}_t}$. Средняя оценка сезонной компоненты для квартала за все годы: $\bar{S}_t = \frac{\sum s_t}{n}$. Скорректированная сезонная компонента: $S_t = \bar{S}_t \cdot k\,;\, k = \frac{4}{\sum_{t=1}^{4} \bar{S}_t}$.

3 шаг. *Устранение сезонной компоненты S.*Разделим каждый уровень исходного временного ряда на скорректированное значение сезонной компоненты. Получим: T*E=Y/S.

4 шаг. *Расчет значений тренда.*Проведем аналитическое выравнивание ряда (T*E) с помощью линейного тренда. Рассчитаем значения T для каждого момента времени по уравнению тренда.

5 шаг. *Расчет значений T+S.* Умножим уровни T на значения сезонной компоненты (S) для соответствующих кварталов.

6 шаг. *Расчет абсолютной ошибки.*Выполним расчет ошибки для каждого уровня ряда по формуле: E=Y/(T*S). Расчет суммы квадратов абсолютных ошибок и ее сравнение с общей суммой квадратов отклонений уровней ряда.

Вопросы и задания для самоконтроля

1. В чем особенность временного ряда?
2. Каковы основные компоненты уровней временного ряда?
3. В чем состоит основная задача эконометрического исследования временного ряда?
4. Как определяется автокорреляция остатков во временных рядах?
5. Какие свойства имеет коэффициент автокорреляции?
6. Как определяется автокорреляционная функция?
7. Что такое коррелограмма? Что выявляют при помощи анализа коррелограммы?
8. Как сформулировать вывод о структуре временного ряда?
9. Какие методы применяются для выявления основной тенденции ряда?

10. В чем суть сглаживания временных рядов?

Задача 1. Имеются следующие данные об урожайности пшеницы y за 12 лет:

y_t	16,3	20,2	17,1	9,7	15,3	16,3	19,9	14,4	18,7	20,7	19,5	21,1
t	1	2	3	4	5	6	7	8	9	10	11	12

Задание:

1) определить среднее значение, среднее квадратическое отклонение и коэффициенты автокорреляции (для лагов $\tau = 1,2$) временного ряда;

2) провести сглаживание исходного временного ряда методом скользящих средних, используя среднюю арифметическую с интервалом сглаживания:

а) $m = 3$;

б) $m = 4$;

3) записать уравнение тренда ряда, полагая, что он линейный, и проверить его значимость на уровне $\alpha = 0,05$.

Задача 2. Данные, отражающие динамику роста доходов y_t на душу населения за восемь лет, приведены в таблице:

Год, t	1	2	3	4	5	6	7	8
y_t	1130	1220	1350	1390	1340	1380	1490	1680

Задание: определить точечный прогноз дохода населения по линейному тренду на 9 год.

Лекция 3(2)

Тема 17. Модели стационарных и нестационарных временных рядов

Вопросы для изучения:

1. Модели стационарных и нестационарных временных рядов, их идентификация.

2.Модель авторегрессии–скользящего среднего (модель ARMA).

3. Авторегрессионная модель проинтегрированного скользящего среднего (модель ARIMA).

Аннотация. Данная тема раскрывает особенности моделей стационарных и нестационарных временных рядов и методы их оценивания.

Ключевые слова. Стационарный процесс, модель авторегрессии, модель Бокса-Дженкинса.

Методические рекомендации по изучению темы

•Тема содержит лекционную часть, где даются общие представления по теме.

•В дополнение к лекции есть презентация, которую необходимо изучить и ответить на вопросы.

•В качестве самостоятельной работы предлагается подготовить реферат в рамках вопросов для изучения.

•Для проверки усвоения темы имеется тест для самоконтроля.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11766

2. Валентинов, В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 328-338.

3.Тихомиров Н. П. Эконометрика: учебник. - М.: Экзамен, серия «Учебник Плехановской академии», 2007, -512 с. С.211-222.

4. Эконометрика: учебник / под ред. В. С. Мхитаряна. - М.: Проспект, 2008. -384 с. С. 325-336.

5. Электронный курс “Time Series Econometrics”, PrincetonUniversity, URL: http://sims.princeton.edu/yftp/Times05/;https://blackboard.princeton.edu/webapps/portal/frameset.jsp?tab_group=courses&url=%2Fwebapps%2Fblackboard%2Fexecute%2FcourseMain%3Fcourse_id%3D_52968_1.

Модели стационарных и нестационарных временных рядов, их идентификация. Набор случайных переменных X(t) называется стохастическим процессом. Стохастический процесс Xt называется стационарным в сильном смысле, если совместное распределение вероятностей всех переменных Xt1, Xt2,…, Xtn, такое же, что и для переменных Xt1+τ, Xt2+τ,…, Xtn+τ. Для стационарного процесса в слабом смысле среднее и дисперсия независимо от рассматриваемого периода времени имеют постоянное значение, а автоковариация зависит только от длины лага между рассматриваемыми переменными.

Временной ряд x1,x2,...,xt, т.е. конкретная реализация стационарного стохастического процесса Xt, также называется стационарным.

Стационарность означает отсутствие:			
тренда	систематических изменений дисперсии	строго периодичных колебаний	систематически изменяющихся взаимозависимостей между элементами временного ряда

Рис. 17.1. Признаки стационарности временного ряда

Линейные модели временных рядов применяются, как правило, для описания стационарных процессов. Чаще всего это стационарные процессы второго порядка, то есть процессы, имеющие постоянные значения всех своих моментов до второго порядка включительно на всех временных отрезках, входящих в интервал $t = 1,2,...,T$. Следовательно, для любых двух интервалов времени $(T_{1,}T_2)$ и $(T_{3,}T_4)$ в таком процессе y_t выполняются условия равенства математических ожиданий, дисперсий и коэффициентов автокорреляции одинаковых порядков. На практике для оценок этих показателей должны выполняться соотношения:

$$\bar{y}_1 = \frac{1}{T_2 - T_1}\sum_{t=T_1}^{T_2} y_t = \frac{1}{T_4 - T_3}\sum_{t=T_3}^{T_4} y_t = \bar{y}_2 \tag{1}$$

$$D_1(y) = \frac{1}{T_2 - T_1}\sum_{t=T_1}^{T_2} (y_t - \bar{y})^2 = \frac{1}{T_4 - T_3}\sum_{t=T_3}^{T_4} (y_t - \bar{y})^2 = D_2(y) \tag{2}$$

$$r_i^{(1)} = \sum_{t=T_1}^{T_2 - i} \frac{(y_t - \bar{y})(y_{t+i} - \bar{y})}{(T_2 - T_1 - i)\cdot D(y)} = \sum_{t=T_3}^{T_4 - i} \frac{(y_t - \bar{y})(y_{t+i} - \bar{y})}{(T_4 - T_3 - i)\cdot D(y)} = r_i^{(2)}, i = 1,2,... \tag{3}$$

Здесь $\bar{y}^{(1)}$, $\bar{y}^{(2)}$, $D^{(1)}(y)$, $D^2(y)$, $r_i^{(1)}$, $r_i^{(2)}$ - оценки математических ожиданий, дисперсий и коэффициентов автокорреляции i - го порядка процесса yt на первом и втором интервалах; $\bar{y}$ и $D(y)$ - оценки среднего значения и дисперсии процесса на интервале $(1,T)$.

Равенства (1) – (3) понимаются в статистическом смысле. Это значит, что каждое из этих равенств может в точности не выполняться, однако разница между оценками укладывается в границы соответствующего критерия.

Такие критерии реализуются через различные тесты, которые мы сейчас рассмотрим.

Параметрические тесты стационарности применяются при достаточно строгих предположениях о законе распределения временного ряда и его параметрах. Они оценивают степень близости эмпирических характеристик временного ряда к их теоретическим аналогам. Для выражений (1) – (3) параметрическими критериями стационарности являются критерии Стьюдента и Фишера. Здесь предполагается нормальный закон распределения значений временного ряда и его выборочных характеристик, что справедливо для многих реальных процессов.

Тестирование математического ожидания по статистике Стьюдента требует разбить временной ряд (1,T) на две части, не обязательно одинаковые, H_0 – гипотеза о постоянстве математического ожидания:

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\sqrt{\frac{s_1^2}{T_1}} + \sqrt{\frac{s_2^2}{T_2}}}}, \sigma_1^2 \neq \sigma_2^2$$

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{s^2} \cdot \sqrt{\frac{T_1 \cdot T_2}{T_1 + T_2}}, \sigma_1^2 = \sigma_2^2 = \sigma$$

$$\tau < \tau(p, v = T_1 + T - 2) \Rightarrow H_0$$

Тестирование математического ожидания по статистике Фишера (если количество наблюдений достаточно велико), H_0 – гипотеза о постоянстве математического ожидания временного ряда. Интервал наблюдений делится на несколько частей.

$$F = \frac{\frac{1}{n-1} \cdot \sum_{j=1}^{n} T_j (\bar{y}_j - \bar{y})^2}{\bar{s}^2(n)}$$

$$\bar{s}^2(n) = \frac{1}{T-n} \cdot \sum_{j=1}^{n} (T_j - 1) \cdot \bar{s}_j^2$$

$$F < F(p, v_1 = n-1, v_2 = T_1 + T_2 + ... + T_n - n \Rightarrow H_0$$

где, n – число частей разбиения интервала (1,T); T_j- число измерений переменной y_t на j-ой части; j=1,2,…,n;

$\bar{y}$ - среднее значение временного ряда;

$\bar{s}^2(n)$ - средняя дисперсия.

Тестирование дисперсии временного ряда на постоянство ее значения проводится разбиением исходного интервала на две части с использованием двухстороннего критерия Фишера, который рассчитывается по формуле:

$$F = \frac{s_1^2}{s_2^2},$$

где s_1^2 и s_2^2 - оценки дисперсии ряда на первой и второй подвыборок соответственно и числом измерений T1 и T2.

Этот тест аналогичен проверке гипотезы о равенстве дисперсий двух нормальных CB с той разницей, что здесь сравнивается дисперсия в разных частях одного временного ряда.

Рассчитанное значение F - статистики сравнивается с критическими на уровнях α/2 и 1-α/2 и числами степеней свободы (Т1-1) и (Т2-1). Если $F_{кр.}(1-\alpha/2;T_1-1;T_2-1) \le F \le F_{кр.}(\alpha/2;T_1-1;T_2-1)$, то гипотеза $\sigma_1^2 = \sigma_2^2 = \sigma^2$ принимается на уровне α.

Поскольку критические значения удовлетворяют соотношению

$$F(\alpha/2;v_1;v_2) = \frac{1}{F(1-\alpha/2;v_1;v_2)},$$

на практике проверяется только соотношение $F \le F_{кр.}(\alpha/2;T_1-1;T_2-1)$

при условии, что $s_1^2 \ge s_2^2$.

При достаточно больших объемах наблюдений временного ряда $(T \ge 40)$ вместо критерия Фишера рекомендуют использовать стандартизированное нормальное распределение. При выборках $40 \le T \le 100$ закону $N(0,1)$ подчиняется случайная величина

$$\varPhi = \frac{\frac{1}{2}\ln\frac{s_1^2}{s_2^2} + \frac{1}{2}\left(\frac{1}{v_1} - \frac{1}{v_2}\right)}{\sqrt{\frac{1}{2}\left(\frac{1}{v_1} + \frac{1}{v_2}\right)}}$$

При больших выборках расчетное значение случайной величины определяется так:

$$\varPhi = (s_1 - s_2)\sqrt{\frac{s_1^2}{2T_1} + \frac{s_2^2}{2T_2}}$$

В любом случае, если $\Phi < \varPhi_{\alpha/2}$, то гипотеза о постоянстве дисперсии принимается.

Если временной ряд разбивается на большее число частей (n>2), гипотеза о постоянстве дисперсии может быть проверена критерием Кокрена, основанном на распределении Фишера. При этом обычно объемы этих частей принимаются равными между собой, то есть Т1=Т2=…=Тn. Рассчитывается критерий по формуле:

$$K = \frac{s_{\max}^2}{s_1^2 + \ldots + s_n^2}$$

Здесь $s_{\max}^2 = \max_j \left(s_j^2\right), \quad j = \overline{1,n}$.

Табличное значение критерия Кокрена определяется по формуле:

$$K(\alpha; n; N-1) = \frac{F(1-\alpha/n; N-1; (n-1)(N-1))}{(n-1) + F(1-\alpha/n; N-1; (n-1)(N-1))}$$

Если $K < K(\alpha; n; N-1)$,то гипотеза о постоянстве дисперсии временного ряда принимается на уровне α.

Критерий Бартлетта также используется при проверке гипотезы о постоянстве дисперсии. Он является более мощным, чем критерий Кокрена, но и более чувствительным к отклонениям значений временного ряда от нормального закона. Здесь временной ряд также разбивается на несколько частей, причем не обязательно одинаковых по величине.

Согласно критерию, следующая величина:

$$\lambda = -\frac{1}{c}\sum_{i=1}^{n}(T_i - 1)\ln\frac{s_i^2}{s^2}$$

распределена приблизительно по закону λ^2 с $(n-1)$ степенями свободы. Здесь

$$s^2 = \frac{\sum_{i=1}^{n}(T_i - 1)s_i^2}{\sum_{i=1}^{n}(T_i - 1)}$$ -

средняя дисперсия на n интервалах;

$$c = 1 + \frac{1}{3(n-1)} \cdot \sum_{i=1}^{n}\frac{1}{T_i - 1} - \frac{1}{\sum_{i=1}^{n}(T_i - 1)}$$

При больших Ti $c \approx 1$.

Для одинаковых размеров подвыборок $v_1 = v_2 = \ldots = v_n = v$, тогда $\sum_{i=1}^{n}(T_i - 1) = T - n$, поэтому $\lambda = \frac{1}{c} nv\left(\ln \bar{s}^2 - \frac{1}{n}\sum_{i=1}^{n} s_i^2\right)$, где $c = 1 + \frac{n+1}{3 \cdot n \cdot v}$.

Если $\lambda < \chi^2(\alpha, n-1)$, то гипотеза о равенстве дисперсий на рассматриваемых частях временного интервала принимается.

По рассмотренным параметрическим критериям следует отметить их ограниченность в применении вследствие достаточно жестких предположений нормальности закона распределения временного ряда. Кроме того, они требуют значительных вычислений. Однако реальные временные ряды могут быть распределены по закону, отличному от нормального. Поэтому на практике при проверке стационарности процессов часто используют непараметрические критерии, не имеющие ограничений по закону распределения и не столь сложные по вычислениям.

Тест Манна – Уитни используется вместо критерия Стьюдента для проверки идентичности распределений двух совокупностей, то есть временных последовательностей одного временного ряда, определенных на разных временных частях интервала 1,2,…,T. Он тестирует постоянство математического ожидания.

Пусть первая подвыборка образована T1 последовательными значениями yt, а вторая – T2 его последовательными значениями, и эти последовательности не пересекаются.

Обозначим элементы первой подвыборки символом $y^{(1)}$, второй – символом $y^{(2)}$. Затем объединим эти подвыборки в одну совокупность объемом (T1+T2), расположив все элементы в порядке возрастания их значений. При этом элементы подвыборок оказываются перемешанными между собой.

Если ряд стационарный, то элементы разных подвыборок довольно равномерно перемешаны друг с другом. В противном случае общая последовательность оказывается разделенной на массивы, состоящие в основном из единиц одной из совокупностей. Например, для возрастающего или убывающего временного ряда элементы подвыборок скапливаются на разных концах общей последовательности.

В тесте Манна – Уитни проверяется гипотеза о стационарности временного ряда на основе критерия u^*, равного числу случаев, когда элементы из первой подвыборки предшествуют элементам из второй подвыборки. Значение u^* рассчитывается по формулам: $u^* = R_1 - \frac{T_1(T_1+1)}{2}$ или

$$u^* = T_1 \cdot T_2 - \frac{T_1(T_1+1)}{2} - R_2,$$

где R_1 и R_2 - суммы рангов элементов первой и второй подвыборок соответственно, определяемых по их общей последовательности.

Для достаточно больших последовательностей (T>50) случайная величина u^* распределена по нормальному закону с математическим ожиданием $M[u^*] \approx \frac{T_1 \cdot T_2}{2}$ и дисперсией $D[u^*] \approx \frac{T_1 \cdot T_2 \cdot (T_1 + T_2 + 1)}{12}$.

Таким образом, случайная величина $z = \frac{u^* - \frac{T_1 \cdot T_2}{2} \pm \frac{1}{2}}{\sigma(u^*)}$ распределена по закону $N(0,1)$. Поправка 1/2 вводится для обеспечения непрерывности z. Она прибавляется, если z<0. Она прибавляется, если z<0, и вычитается, если z>0

Если обе подвыборки идентичны, их элементы перемешаны между собой, тогда значение u^* будет находиться вблизи своего среднего значения, а значение z - около нуля. Поэтому гипотеза о стационарности процесса yt принимается на уровне значимости α, если выполняется неравенство:

$$u_{1-\alpha/2} \le z \le u_{\alpha/2}$$

Непараметрический тест Сигела – Тьюки используют вместо параметрического критерия Фишера для проверки гипотезы о постоянстве дисперсии временного ряда. Он также основан на сопоставлении рангов элементов двух подвыборок из данного интервала.

Сначала исходный временный ряд центрируется, то есть каждое значение заменяется отклонением от среднего согласно выражению $y_{\mathring{a}} = y_t - \bar{y}$, где $\bar{y}$ - среднее значение ряда yt.

Далее интервал (1,T) делится на две, желательно равные, части, где элементы обозначаются соответственно $y^{(1)}$ и $y^{(2)}$. Эти элементы в объединенной совокупности сортируются в порядке возрастания их значений. Затем каждому значению присваивается его ранг по следующему правилу: все нечетные номера получают отрицательные элементы в порядке возрастания их значений, а все четные номера – положительные элементы, но в порядке убывания их значений. Другими словами, ранг 1 получает наименьшее отрицательное значение, а ранг 2 – наибольшее положительное.

Если обозначить R_1 сумму рангов элементов первой подвыборки, то случайная величина

$$z = \frac{R_1 - \frac{T_1 \cdot (T_1 + T_2 + 1)}{2} \pm \frac{1}{2}}{\sqrt{\frac{T_1 \cdot T_2 \cdot (T_1 + T_2 + 1)}{12}}}$$

оказывается распределенной по закону $N(0,1)$. Здесь также поправка 1/2 вводится для обеспечения непрерывности z.

Отдельную группу непараметрических тестов стационарности составляют тесты, основанные на так называемых сериальных критериях. Они анализируют закономерности серий измеренных значений временного ряда. Для их применения необходим достаточно большой объем данных, чтобы считать обнаруженные закономерности устойчивыми.

Серией называют последовательность значений временного ряда, отклоняющихся от значения некоторого признака в одну и ту же сторону. Например, при тестировании автокорреляции в остатках по методу рядов таким признаком было расчетное значение результативного признака. Во временных рядах в роли этого признака часто выступает медиана значений ряда. Тогда элементы, по значению превышающие медиану, образуют серии с положительным знаком, а элементы, не превосходящие по значению медиану – серии с отрицательным знаком.

Критерий Вальда – Вольфовица основан на подсчете общего числа серий. Среднее число серий рассчитывается по выражению:

$$M\left[N_s\right]=\frac{2N_1N_2}{N_1+N_2}+1$$

, а дисперсия – по формуле:

$$D\left[N_s\right]=\frac{2N_1N_2\cdot\left[2N_1\cdot N_2-\left(N_1+N_2\right)\right]}{\left(N_1+N_2\right)^2\cdot\left(N_1+N_2-1\right)}$$

Здесь N1 и N2 - количества элементов соответственно с положительным и с отрицательным знаком; (N1+N2)=T; N_s - число серий.

Как видим, эти формулы в точности повторяют формулы метода рядов.

При большом объеме временного ряда случайная величина

$$z=\frac{N_s-M\left[N_s\right]\pm\frac{1}{2}}{\sigma\left(N_s\right)}$$

распределена по закону $N(0;1)$.

Если реальный временной ряд не представляет стационарный процесс второго порядка, его нужно привести к стационарному процессу. Это делается с помощью соответствующих преобразований: взятия конечных разностей, логарифмирования цепных индексов, расчета темпов прироста и др.

Например, когда закон изменения y_t близок к линейному, преобразование заключается во взятии первых разностей:

$$y_t'=\Delta y_t=y_t-y_{t-1}$$

Разности второго порядка:

$$y_t''=\Delta y_t'=y_t'-y_{t-1}'=\left(y_t-y_{t-1}\right)-\left(y_{t-1}-y_{t-2}\right)=y_t-2y_{t-1}+y_{t-2}$$

применяются при законе изменения y_t, близком к квадратической параболе.

При экспоненциальном росте y_t логарифмируются цепные индексы:

$$y_t'=\ln\frac{y_t}{y_{t-1}}=\ln y_t-\ln y_{t-1}$$

Расчет темпов прироста выполняется по формуле

$$y'_t = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1$$

Для трансформации исходного нестационарного ряда в стационарный можно использовать и другие преобразования. В каждом конкретном случае надо исходить из примерной формы временного графика y_t. Подходящее преобразование должно обеспечивать приблизительное выполнение условия $y'_t = f(y_t) = const$.

Особенности конкретного стационарного процесса второго порядка полностью определяются характером его автокорреляционной функции, представляющей собой последовательность коэффициентов автокорреляции r_0, r_1, r_2, … Здесь $r_0 = 1$, остальные значения располагаются на отрезке $[-1;1]$.

Аналогично формируется автокорреляционная функция как последовательность значений автокорреляций γ_0, γ_1, γ_2,… в зависимости от сдвига. Между значениями двух функций существует взаимосвязь $\gamma_i = r_i \sigma_y^2$, i=0,1,2,…; $\gamma_0 = \sigma_y^2$.

Модель авторегрессии–скользящего среднего (модель ARMA). Построение модели АР(k) сводится к решению двух задач:

- определение рационального порядка модели (величины k);
- оценивание параметров модели на основе уравнений Юла-Уокера.

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_k y_{t-k} + \varepsilon_t$$

Система уравнений Юла-Уокера:

$$r_1 = a_1 + a_2 r_1 + ... + a_k r_{k-1};$$
$$r_2 = a_1 r_1 + a_2 + ... + a_k r_{k-2};$$
$$\cdots\cdots\cdots$$
$$r_k = a_1 r_{k-1} + a_2 r_{k-2} + ... + a_k;$$

$r_1, r_2, \ldots r_k$ – известные оценки коэффициентов автокорреляции;

$a_1, a_2, \ldots a_k$ - неизвестные оценки коэффициентов модели.

Модель авторегрессии первого порядка АР(1):

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t,$$
$$a_1 = r_1$$

Модель авторегрессии второго порядка АР(2):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t,$$
$$r_1 = a_1 + a_2 r_1;$$
$$r_2 = a_1 r_1 + a_2.$$
$$a_1 = \frac{r_1(1-r_2)}{1-r_1^2};$$
$$a_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

Модель скользящего среднего первого порядка СС(1):

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1},$$
$$\sigma_y^2 = (1+\beta_1^2)\sigma_\varepsilon^2,$$
$$\rho_1 = \frac{-\beta_1}{1+\beta_1^2}$$

Модель скользящего среднего второго порядка СС(2):

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2},$$
$$\sigma_y^2 = (1+\beta_1^2+\beta_2^2)\sigma_\varepsilon^2,$$
$$\rho_1 = \frac{-\beta_1(1-\beta_1)}{1+\beta_1^2+\beta_2^2};\ \rho_2 = \frac{-\beta_2}{1+\beta_1^2+\beta_2^2};\ \rho_i = 0, i \geq 3.$$

Простейшая модель авторегрессии - скользящего среднего АРСС(k,m) - (**A**uto**R**egressive-**M**oving**A**verage (ARMA (k,m)) :

$$y_t = \alpha y_{t-1} + \varepsilon_t - \beta \varepsilon_{t-1}$$
$$y_t - \alpha y_{t-1} = \varepsilon_t - \beta \varepsilon_{t-1}, |\alpha| < 1, |\beta| < 1$$

Значения автокорреляционной функции для ARMA (1,1) будут иметь вид:

$$\rho(1) = \frac{(1-\alpha\beta)(\alpha-\beta)}{1+\beta^2-2\alpha\beta}, \tau = 1$$
$$\rho(\tau) = \alpha\rho(\tau-1) = \alpha^{\tau-1}\rho(1), \tau > 1$$

Авторегрессионная модель проинтегрированного скользящего среднего (модель ARIMA). Для описания нестационарных однородных временных рядов применяется модель Бокса-Дженкинса (ARIMA –модель). Наиболее распространены ARIMA (k,m,q) – модели, со значениями параметров, не превышающими 2, q – порядок разности (дискретной производной).

Этапы методологии Бокса-Дженкинса:

1. Тестирование исходного ряда на стационарность. Анализ автокорреляционной функции. Переход к стационарному ряду путем взятия последовательных разностей (дискретные производные). Определение параметра q.

2. Исследование характера автокорреляционной функции и предположение о значениях параметров k (порядок авторегрессии) и m (порядок скользящего среднего).

3. Оценивание параметров ARIMA (k,m,q) – модели.

4. Проверка пробной модели на адекватность путем анализа ряда остатков.

Для обнаружения «белого шума» в остатках применяют Q-статистику Бокса-Пирса, H_0 об отсутствии автокорреляции в остатках:

$$Q = n\sum_{p=1}^{\tau} r_p^2,$$

$$Q < \chi^2(\alpha, v = \tau - k - m) \Rightarrow H_0 : \rho = 0$$

Критерии качества подгонки модели Бокса-Дженкинса:

КритерийАкайка (Akaike information criterion, AIC):

$$AIC = \frac{k+m}{n} + \ln\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right)$$

Выбор следует сделать в пользу модели с меньшим значением AIC.

Критерий Шварца (Swarzcriterion):

$$SIK = \frac{(p+q)\ln n}{n} + \ln\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right)$$

Вопросы для самоконтроля

1. Какая модель временного ряда называется статической?

2. Когда модель временного ряда называется динамической?
3. Как определяются авторегрессионные модели?
4. Как определяется модель ARMA?
5. Как интерпретируют параметры моделей авторегрессии?
6. Что означает стационарность временного ряда?
7. Какой стационарный процесс называется «белым шумом»?
8. Какие типы включают модели стационарных временных рядов?
9. Какие типы включают модели нестационарных временных рядов?
10. Как определяется ARIMA-модель?

Лекция 4(1)

Тема 19. Понятие о системах эконометрических уравнений

Вопросы для изучения:

1. Понятие о системах уравнений. Системы независимых уравнений и системы взаимозависимых уравнений.

2. Структурная и приведенная формы модели.

3. Идентификация модели.

Аннотация. Данная тема излагает типы систем эконометрических уравнений.

Ключевые слова. Система взаимозависимых уравнений, идентификация системы взаимозависимых уравнений, структурная и приведенная формы модели.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где даются общие представления по теме.
- В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы.
- В качестве самостоятельной работы предлагается выполнить практические задания.

•Для проверки усвоения темы имеется тест для самоконтроля.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11766

2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none)С. 117-136.

3. Валентинов В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 338-356.

4.Эконометрика. Практикум: [Электронный ресурс] Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов.знание, 2014. - 329 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 286-313.

Понятие о системах уравнений. Системы независимых уравнений и системы взаимозависимых уравнений. Объектом статистического изучения в социальных науках являются сложные системы. Построение изолированных уравнений регрессии недостаточно для описания таких систем и объяснения механизма их функционирования. Изменение одной переменной, как правило, не может происходить без изменения других. Поэтому важное место занимает проблема описания структуры связей между переменными системой так называемых одновременных уравнений. Так, если изучается модель спроса как отношение цен и количества потребляемых товаров, то одновременно для прогнозирования спроса необходима модель предложения товаров, в которой рассматривается также взаимосвязь между количеством и ценой предлагаемых благ. Это позволяет достичь равновесия между спросом и предложением.

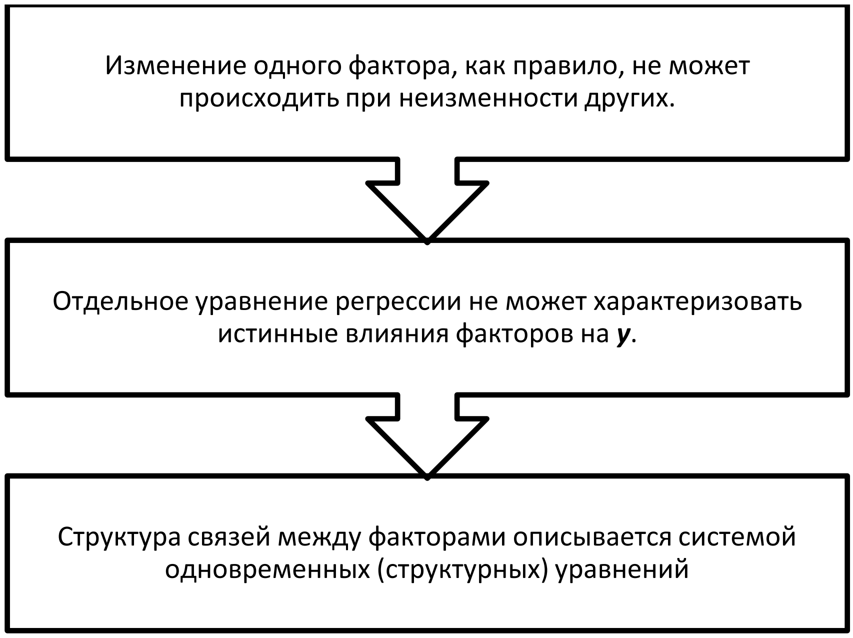


Рис. 19.1. Необходимость систем уравнений

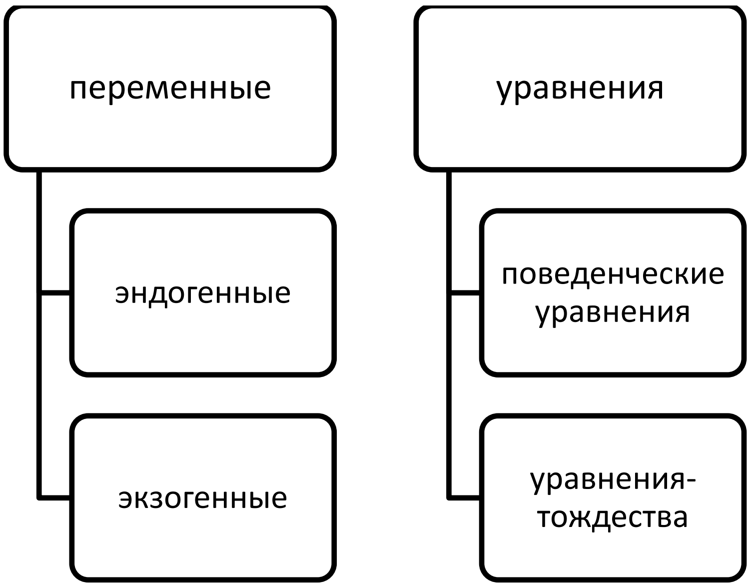


Рис. 19.2. Составляющие систем уравнений

Эндогенные переменные обычно обозначаются как y. Это зависимые переменные, значения которых определяются внутри модели. Их число равно числу уравнений в системе.

Экзогенные переменные обычно обозначаются как x. Это внешние по отношению к модели переменные. Они влияют на эндогенные переменные, но не зависят от них.

Лаговые переменные – это значения эндогенных переменных за предшествующий период времени (yt-1). В модели участвуют в качестве экзогенных переменных.

В поведенческих уравнениях описываются взаимодействия между переменными.

В уравнениях-тождествах описываются соотношения, которые должны выполняться во всех случаях. Тождества не содержат подлежащие оценке параметры a и b, а также случайное отклонение ε.

Рис. 19.3. Виды систем уравнений

В системе независимых уравнений каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x:

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

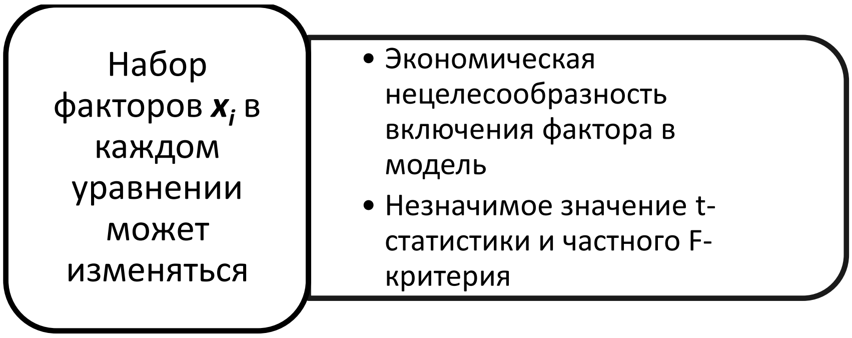


Рис. 19.4. Включение факторов в модель

Система независимых уравнений с различным набором факторов:

$$\begin{cases} y_1 = f(x_1, x_2, x_3, x_4, x_5), \\ y_2 = f(x_1, x_3, x_4, x_5), \\ y_3 = f(x_2, x_3, x_5), \\ y_4 = f(x_3, x_4, x_5). \end{cases}$$

В системе рекурсивных уравнений каждое последующее уравнение включает в качестве факторов все зависимые переменные у предшествующих уравнений наряду с набором собственно факторов х:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3m}x_m + \varepsilon_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + b_{n3}y_3 + \ldots + b_{nn-1}y_{n-1} + \\ + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

В системе взаимозависимых уравнений одни и те же зависимые переменные у в одних уравнениях входят в левую часть, а в других уравнениях – в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

Этот вид систем уравнений получил наибольшее распространение в эконометрических исследованиях. В эконометрике эта система уравнений называется также структурной формой модели (СФМ). Для нахождения параметров каждого уравнения традиционный МНК неприменим, здесь используются специальные методы оценивания. В этом случае каждое из уравнений не может рассматриваться самостоятельно.

Структурная и приведенная формы модели. Система взаимозависимых (одновременных) уравнений, описывающая структуру связей между переменными, называется структурной формой модели. Коэффициенты b_i и a_j называются структурными коэффициентами модели. Приведенная форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных. В каждое приведенное уравнение включаются все экзогенные переменные структурной модели. Система одновременных уравнений (т.е. структурная форма модели) обычно содержит эндогенные и экзогенные переменные. Эндогенные переменные – это зависимые переменные, число которых равно числу уравнений в системе. Они обозначаются через у. Экзогенные переменные – это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них. Они обозначаются через x.

Простейшая структурная форма модели имеет вид:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_{1,} \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_{2,} \end{cases}$$

где y_1,y_2 – эндогенные переменные, x_1,x_2 – экзогенные.

Классификация переменных на эндогенные и экзогенные зависит от теоретической концепции принятой модели. Экономические переменные могут выступать в одних моделях как эндогенные, а в других – как экзогенные переменные. Внеэкономические переменные (например, климатические условия) входят в систему как экзогенные переменные. В качестве экзогенных переменных можно рассматривать значения эндогенных переменных за предшествующий период времени (лаговые переменные). Например, потребление текущего года yt может зависеть также и от уровня потребления в предыдущем году y_{t-1}.

Структурная форма модели позволяет увидеть влияние изменений любой экзогенной переменной на значения эндогенной переменной. Целесообразно в качестве экзогенных переменных выбирать такие переменные, которые могут быть объектом регулирования. Меняя их и управляя ими, можно заранее иметь целевые значения эндогенных переменных.

Коэффициенты b_i при эндогенных и a_j – при экзогенных переменных называются структурными коэффициентами модели. Все переменные в модели могут быть выражены в отклонениях $(x-\bar{x})$ и $(y-\bar{y})$ от среднего уровня, и тогда свободный член в каждом уравнении отсутствует.

Использование МНК для оценивания структурных коэффициентов модели дает смещенные и несостоятельные оценки. Поэтому обычно для определения структурных коэффициентов модели структурная форма преобразуется в приведенную.

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

$$\begin{cases} \hat{y}_1 = \delta_{11}x_1 + \ldots + \delta_{1m}x_m, \\ \hat{y}_2 = \delta_{21}x_1 + \ldots + \delta_{2m}x_m, \\ \ldots \\ \hat{y}_n = \delta_{n1}x_1 + \ldots + \delta_{nm}x_m. \end{cases}$$

δ_{ij} – коэффициенты приведенной формы модели.

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

Приведенная форма позволяет выразить значения эндогенных переменных через экзогенные, однако аналитически уступает структурной форме модели, т.к. в ней отсутствуют оценки взаимосвязи между эндогенными переменными.

Идентификация модели. При переходе от приведенной формы модели к структурной исследователь сталкивается с проблемой идентификации. Идентификация – это единственность соответствия между приведенной и структурной формами модели.

Структурная модель в полном виде, состоящая в каждом уравнении системы из n эндогенных и m экзогенных переменных, содержит n(n-1+m) параметров. Приведенная модель в полном виде содержит nm параметров. Таким образом, в полном виде структурная модель содержит большее число параметров, чем приведенная форма модели. Поэтому n(n-1+m) параметров структурной модели не могут быть однозначно определены через nm параметров приведенной формы модели.

Чтобы получить единственно возможное решение для структурной модели, необходимо предположить, что некоторые из структурных коэффициентов модели равны нулю. Тем самым уменьшится число структурных коэффициентов.

С позиции идентифицируемости структурные модели можно подразделить на три вида: идентифицируемые; неидентифицируемые; сверхидентифицируемые.

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

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

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

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Если же в системе нет неидентифицируемых уравнений и имеется хотя бы одно сверхидентифицируемое, то модель будет сверхидентифицируемой.

Обозначим H – число эндогенных переменных в i- ом уравнении системы, D – число экзогенных переменных, которые содержатся в системе, но не входят в данное уравнение. Тогда условие идентифицируемости уравнения может быть записано в виде следующего счетного правила:

$D+1 = H$ – уравнение идентифицируемо;

$D+1 < H$ – уравнение неидентифицируемо;

$D+1 > H$ – уравнение сверхидентифицируемо.

Это счетное правило отражает необходимое, но не достаточное условие идентификации. Достаточное условие идентификации отдельного уравнения состоит в том, чтобы матрица из коэффициентов при переменных, которые в данном уравнении отсутствуют (то есть коэффициенты берутся из всех остальных уравнений системы), имела ранг не меньший, чем количество эндогенных переменных в системе минус единица.

Следует помнить, что на идентификацию проверяется каждое уравнение модели.

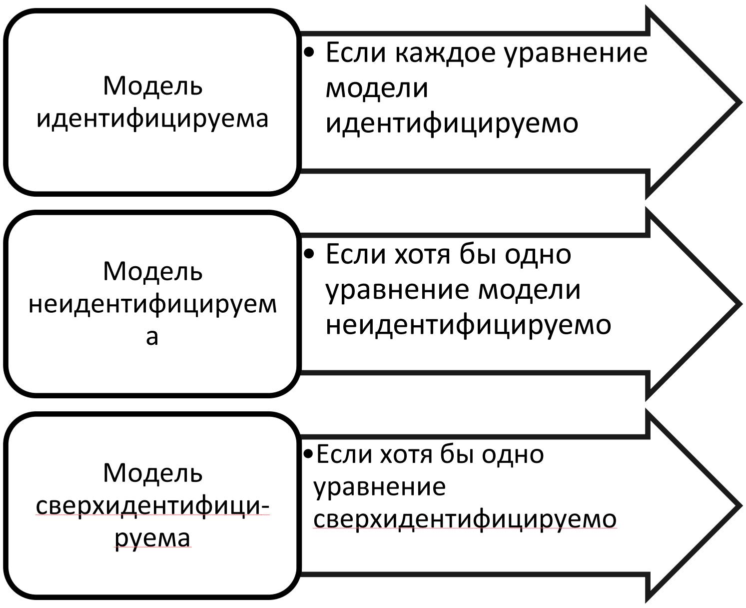


Рис. 19.5. Вывод об идентификации

Вопросы и задания для самоконтроля

1. В чем преимущество систем эконометрических уравнений?
2. Какие переменные называют предопределенными?
3. Что такое структурная форма модели?
4. Что такое приведенная форма модели?
5. Почему нужна приведенная форма модели?
6. Когда структурная модель является идентифицируемой?
7. Когда структурная модель является неидентифицируемой?
8. В каком случае модель является сверхидентифицируемой?
9. Как идентифицируется отдельное уравнение в системе по счетному правилу?
10. В чем состоит достаточное условие идентификации отдельного уравнения?

Задача 1. Дана модель Менгеса:

$$Y_t = \alpha_1 + b_{11} Y_{t-1} + b_{12} I_t + \varepsilon_1,$$

$$I_t = \alpha_2 + b_{21} Y_t + b_{22} Q_t + \varepsilon_2,$$

$$C_t = \alpha_3 + b_{13} Y_t + b_{32} C_{t-1} + b_{33} P_t + \varepsilon_3,$$

$$Q_t = \alpha_4 + b_{41} Q_{t-1} + b_{42} R_t + \varepsilon_4.$$

где Y - национальный доход; C - расходы на личное потребление; I - чистые инвестиции; Q - валовая прибыль экономики; P - индекс стоимости жизни; R - объем продукции промышленности; t - текущий период; $(t-1)$ - предыдущий период.

Задание: проверить идентифицируемость каждого уравнения с использованием необходимого и достаточного условий идентификации.

Задача 2. Имеется модель денежного и товарного рынков:

$$R_t = \alpha_1 + b_{12}Y_t + b_{14}M_t + \varepsilon_1,$$

$$Y_t = \alpha_2 + b_{21}R_t + b_{23}I_t + b_{25}G_t + \varepsilon_2,$$

$$I_t = \alpha_3 + b_{31}R_t + \varepsilon_3,$$

где R - процентные ставки; Y - реальный ВВП; M - денежная масса; I - внутренние инвестиции; G - реальные государственные расходы; t - текущий период.

Задание: проверить идентифицируемость каждого уравнения с использованием необходимого и достаточного условий идентифицируемости и записать приведенную форму модели.

Лекция 4(2)

Тема 20.Методы оценки систем одновременных уравнений

Вопросы для изучения:

1. Косвенный, двухшаговый и трехшаговый МНК.

2. Применение систем уравнений для построения макроэкономических моделей и моделей спроса – предложения.

Аннотация. Данная тема раскрывает методы оценки систем эконометрических уравнений.

Ключевые слова. Косвенный метод наименьших квадратов, двухшаговый метод наименьших квадратов, модели спроса-предложения.

Методические рекомендации по изучению темы

•Тема содержит лекционную часть, где даются общие представления по теме.

•В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы.

•В качестве самостоятельной работы предлагается выполнить практические задания.

•Для проверки усвоения темы имеется тест.

Рекомендуемые информационные ресурсы:

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11766

2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none) С. 117-136.

3. Валентинов В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none) С. 338-356.

4.Эконометрика. Практикум: [Электронный ресурс] Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов.знание, 2014. - 329 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none) С. 286-313.

Косвенный, двухшаговый и трехшаговый МНК. Косвенный МНК применяется для оценивания идентифицируемых систем одновременных уравнений.

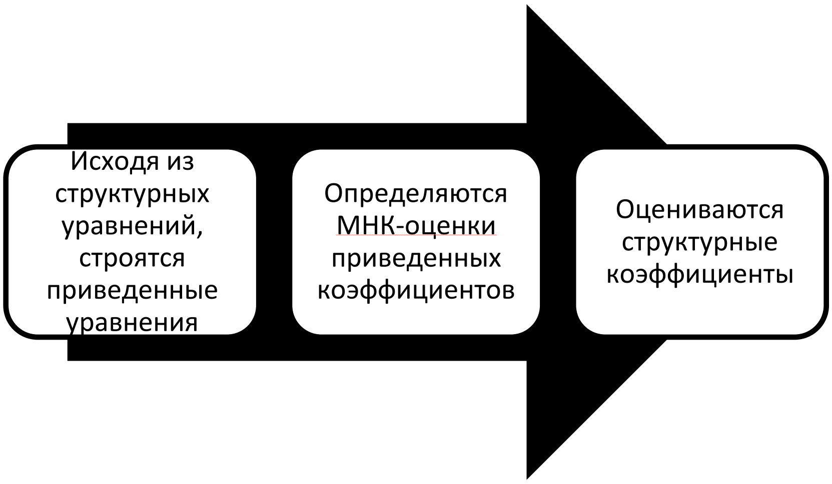


Рис. 20.1. Этапы косвенного МНК

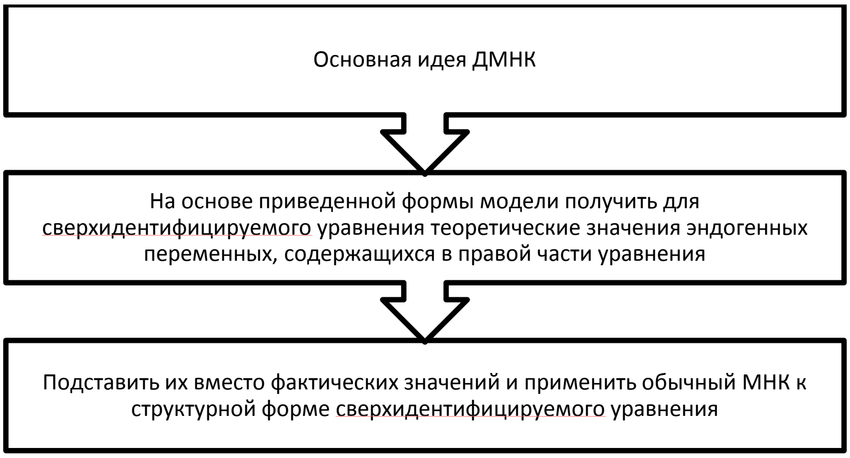


Рис. 20.2. Основная идея двухшагового метода наименьших квадратов

Таким образом, метод наименьших квадратов используется дважды:

1) При определении приведенной формы модели и нахождении на ее основе оценок эндогенной переменной $\hat{y}_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \ldots + \delta_{ij}x_j$

2) При определении структурных коэффициентов структурного сверхидентифицируемого уравнения на основе оценок эндогенных переменных.

Двухшаговый метод наименьших квадратов является наиболее общим и широко распространенным методом решения системы одновременных уравнений. Для точно идентифицируемых уравнений ДМНК дает тот же результат, что и КМНК.

Трехшаговый МНК разработан для оценки одновременно всех уравнений структурной формы модели с учетом возможной взаимной коррелированности регрессионных остатков различных уравнений системы. Этот метод оказывается более эффективным, чем ДМНК, если случайные остатки различных уравнений системы взаимно коррелированы, т.е. если их взаимная ковариационная матрица отлична от диагональной. Однако и в этой ситуации ДМНК – оценки структурных параметров системы остаются состоятельными.

В трехшаговом МНК сохранены первые два шага ДМНК. Однако полученные в результате этих двух шагов, отдельно для каждого уравнения, оценки структурных параметров не являются окончательными, а пересчитываются на 3 – м шаге следующим образом. Оценки структурных коэффициентов используются для подсчета выборочной ковариационной матрицы случайных остатков. Последняя, в свою очередь, используется для одновременного вычисления оценок всех структурных параметров системы с помощью обобщенного МНК в рамках соответствующим образом построенной обобщенной линейной модели множественной регрессии.

Применение систем уравнений для построения макроэкономических моделей и моделей спроса – предложения

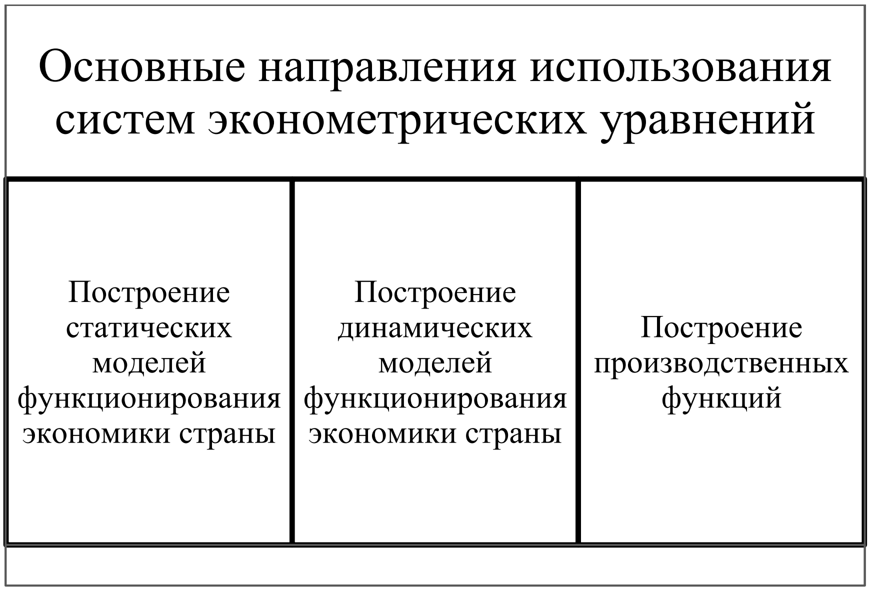


Рис. 20. 3. Основные направления использования систем эконометрических уравнений

Наиболее широко системы одновременных уравнений используются при построении макроэкономических моделей экономики страны. В большинстве случаев это мультипликаторные модели кейнсианского типа. Статическая модель Кейнса в самом простом виде следующая:

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

где C – личное потребление;

y – национальный доход в постоянных ценах;

I – инвестиции в постоянных ценах.

В силу наличия тождества в модели (второе уравнение системы) $b \le 1.$ Он характеризует предельную склонность к потреблению. Если $b = 0{,}65,$ из каждой дополнительной тысячи рублей дохода на потребление расходуется в среднем 650 рублей и 350 рублей инвестируется. Если b>1 то y<C+I, и на потребление расходуются не только доходы, но и сбережения. Параметр a Кейнс истолковывал как прирост потребления за счет других факторов.

Структурный коэффициент b используется для расчета мультипликаторов. По данной функции потребления можно определить два мультипликатора – инвестиционный мультипликатор потребления Mc и национального дохода My :

$M_c = \dfrac{b}{1-b},$ т.е. при $b = 0{,}65$; $M_c = \dfrac{0{,}65}{1-0{,}65} = 1{,}857$.

Это означает, что дополнительные вложения 1 тыс. руб. приведут при прочих равных условиях к дополнительному увеличению потребления на 1,857 тыс. руб.

$M_y = \dfrac{1}{1-b},$ т.е. при $b = 0{,}65$ $M_y = \dfrac{1}{1-0{,}65} = 2{,}857$,

т.е. дополнительные вложения 1 тыс. руб. на длительный срок приведут при прочих равных условиях к дополнительному доходу 2,857 тыс. руб.

Эта модель точно идентифицируема, и для получения b применяется КМНК. Строится система приведенных уравнений:

$$\begin{cases} C = A + B \cdot I + U_1 \\ y = A' + B' \cdot I + U_{2,} \end{cases}$$

в которой $A = A',$ а параметры B и B' являются мультипликаторами, т.е. $B = M_c$ и $B' = M_y$. Для проверки подставим балансовое равенство в первое уравнение структурной модели:

$$C = a + by + \varepsilon = a + b(C + I) + \varepsilon = a + bc + bI + \varepsilon \Rightarrow$$

$$\Rightarrow C(1-b) = a + bI + \varepsilon \Rightarrow C = \underbrace{\frac{a}{1-b}}_{A} + \underbrace{\frac{b}{1-b}}_{M_c} I + \varepsilon \underbrace{\frac{1}{1-b}}_{U_1}$$

Аналогично поступим и со вторым уравнением структурной модели:

$$y = C + I \Rightarrow y = a + by + \varepsilon + I \Rightarrow y(1-b) = a + I + \varepsilon \Rightarrow$$

$$\Rightarrow y = \underbrace{\frac{a}{1-b}}_{A' = A} + \underbrace{\frac{1}{1-b}}_{M_y} I + \underbrace{\frac{1}{1-b}}_{U_2} \varepsilon$$

Таким образом, приведенная форма содержит мультипликаторы, интерпретируемые как коэффициенты множественной регрессии, отвечающие на вопрос, на сколько единиц изменится значение эндогенной переменной, если экзогенная изменится на 1 единицу. Это делает модель удобной для прогнозирования.

В более поздних исследованиях статическая модель Кейнса включала уже не только функцию потребления, но и функцию сбережений:

$$\begin{cases} C = a + by + \varepsilon_{1,} \\ r = T + K(C + I) + \varepsilon_{2,} \\ y = C + I - r, \end{cases}$$

где r – сбережения.

Здесь три эндогенные переменные – C, r и y и одна экзогенная –I. Система идентифицируема: в первом уравнении H=2 и D=2, во втором H=1, D=0; $C + I$ рассматривается как предопределенная переменная.

Наряду со статическими широкое распространение получили динамические модели экономики. Они содержат в правой части лаговые переменные, а также учитывают тенденцию. Например, модель Кейнса экономики США 1950-1960 гг. в упрощенном варианте:

$$\begin{cases} C_t = b_1 S_t + b_2 P_t + b_3 + \varepsilon_{1,} \\ I_t = b_4 P_t + b_5 P_{t-1} + b_6 + \varepsilon_{2,} \\ S_t = b_7 R_t + b_8 R_{t-1} + b_9 t + b_{10} + \varepsilon_{3,} \\ R_t = S_t + P_t + T_{t,} \\ R_t = C_t + I_t + G_t \end{cases}$$

где $T_t -$ чистые трансферты в пользу администрации;

$I_t -$ кап. вложения;

$G_t -$ правительственные расходы;

$S_t -$ заработная плата в период t;

$P_t -$ прибыль;

$P_{t-1} -$ прибыль в период $t-1$;

$R_t -$ общий доход.

Модель содержит 5 эндогенных переменных – C_t, I_t, S_t, R_t (в левой части системы) и P_t (зависимая переменная, определяемая по первому тождеству), три экзогенные переменные – T_t, G_t, t и две лаговые предопределенные переменные P_{t-1} и $R_{t-1.}$ Данная модель сверхидентифицируема и решается ДМНК. Для прогнозных целей используется приведенная форма модели:

$$\begin{cases} C_t = d_1 T + d_2 G + d_3 t + d_4 P_{t-1} + d_5 R_{t-1} + U_{1,} \\ I_t = d_6 T + d_7 G + d_8 t + d_9 P_{t-1} + d_{10} R_{t-1} + U_{2,} \\ S_t = d_{11} T + d_{12} G + d_{13} t + d_{14} P_{t-1} + d_{15} R_{t-1} + U_{3,} \\ R_t = d_{16} T + d_{17} G + d_{18} t + d_{19} P_{t-1} + d_{20} R_{t-1} + U_{4,} \\ P_t = d_{21} T + d_{22} G + d_{23} t + d_{24} P_{t-1} + d_{25} R_{t-1} + U_5 \end{cases}$$

Здесь мультипликаторами являются коэффициенты при экзогенных переменных. Они отражают влияние экзогенной переменной на эндогенную переменную.

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

$$\begin{cases} Q^d = a_0 + a_1 P + \varepsilon_{1,} \text{ - объём спроса,} \\ Q^s = b_0 + b_1 P + \varepsilon_{2,} \text{ - объём предложения,} \\ Q^d = Q^s \end{cases}$$

Здесь 3 эндогенные переменные: Q^d, Q^s и P. При этом, если Q^d и Q^s представляют собой эндогенные переменные, исходя из структуры самой системы, то P является эндогенной по экономическому содержанию (цена зависит от спроса и предложения), а также в результате наличия тождества $Q^d = Q^s$. Приравняем уравнения, получим:

$$a_0 + a_1 P + \varepsilon_1 = b_0 + b_1 P + \varepsilon_2,$$

$$P = \frac{b_0 - a_0}{a_1 - b_1} + \frac{\varepsilon_2 - \varepsilon_1}{a_1 - b_1}.$$

Модель не содержит экзогенной переменной. Однако, чтобы модель имела статистическое решение и можно было убедиться в ее справедливости, в модель вводятся экзогенные переменные.

Например, модель вида:

$$\begin{cases} Q^d = a_0 + a_1 P + a_2 R + \varepsilon_{1,} \\ Q^s = b_0 + b_1 P + b_2 W + \varepsilon_{2,} \\ Q^d = Q^s, \end{cases}$$

где $R-$доход на душу населения; $W-$ климатические условия (при спросе и предложении зерна).

Переменные R и W экзогенные. Введя их в модель, получаем идентифицированную структурную модель.

Динамическая модель Клейна:

$$\begin{cases} C_t = b_1 \cdot S_t + b_2 \cdot P_t + b_3 + \varepsilon_1, \\ I_t = b_4 \cdot P_t + b_5 \cdot P_{t-1} + b_6 + \varepsilon_2, \\ S_t = b_7 \cdot R_t + b_8 \cdot R_{t-1} + b_9 \cdot t + b_{10} + \varepsilon_3, \\ R_t = S_t + P_t + T_t, \\ R_t = C_t + I_t + G_t. \end{cases}$$

Ct – функция потребления в период t; St – заработная плата в период t;Pt – прибыль в период t; Pt-1 – прибыль в период t-1; Rt – общий доход в период t; Rt-1 – общий доход в предыдущий период;t – время;Tt – чистые трансферты в пользу администрации в период t; It – капиталовложения в период t;Gt – спрос административного аппарата, правительственные расходы в период t.

Динамическая модель Клейна сверхидентифицируема и решаема ДМНК. Для прогнозных целей используется приведенная форма модели:

$$\begin{cases} C_t = d_1 T + d_2 G + d_3 t + d_4 P_{t-1} + d_5 R_{t-1} + u_1, \\ I_t = d_6 T + d_7 G + d_8 t + d_9 P_{t-1} + d_{10} R_{t-1} + u_2, \\ S_t = d_{11} T + d_{12} G + d_{13} t + d_{14} P_{t-1} + d_{15} R_{t-1} + u_3, \\ R_t = d_{16} T + d_{17} G + d_{18} t + d_{19} P_{t-1} + d_{20} R_{t-1} + u_4, \\ P_t = d_{21} T + d_{22} G + d_{23} t + d_{24} P_{t-1} + d_{25} R_{t-1} + u_5, \end{cases}$$

В данной модели коэффициенты при экзогенных переменных T и G являются мультипликаторами, отвечающими на вопрос: На сколько единиц изменится значение эндогенной переменной, если экзогенная переменная изменится на одну единицу своего измерения. Коэффициенты d1, d6, d11, d16, d21 – мультипликаторы чистых трансфертов в пользу администрации относительно личного потребления d1, инвестиций d6, заработной платы d11, дохода d16 и прибыли d21. Соответственно коэффициенты d2, d7, d12, d17, d22 являются мультипликаторами правительственных расходов относительно соответствующих эндогенных переменных.

Динамическая модель Кейнса:

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

Yt – имеющийся в распоряжении доход в период времени t; Ct – частное потребление в период времени t; Pt – валовой национальный продукт в период времени t; Yt-1 – доход предыдущего года; Gt - общественное потребление; It – валовые капиталовложения; Lt – изменение складских запасов; Zt – сальдо платежного баланса.

Первое уравнение динамической модели Кейнса является сверхидентифицируемым, а второе и третье – тождествами, доход является лаговой переменной.

Линейная модель спроса и предложения:

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

Qd – спрашиваемое количество благ (объем спроса);

Qs – предлагаемое количество благ (объем предложения).

В этой системе три эндогенные переменные – Qd, Qs и P. При этом если Qd и Qs представляют собой эндогенные переменные исходя из структуры самой системы (они расположены в левой части), то P является эндогенной по экономическому содержанию (цена зависит от предлагаемого и испрашиваемого количества благ), а также в результате наличия тождества Qd=Qs. Линейная модель спроса и предложения не содержит экзогенной переменной. Однако для того, чтобы модель имела статистическое решение и можно было убедиться в ее справедливости, в модель вводятся экзогенные переменные: R и W, после чего модель становится идентифицируемой и может быть оценена КМНК.

$$\begin{cases} Q_d = a_0 + a_1 P + a_2 R + \varepsilon_1, \\ Q_s = b_0 + b_1 P + b_2 W + \varepsilon_2, \\ Q_d = Q_s. \end{cases}$$

R – доход на душу населения;

W – климатические условия (например, спрос и предложение на зерно).

Общий перечень информационных ресурсов

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=382

2. Эконометрика: [Электронный ресурс] Учеб.пособие / А.И. Новиков. - 2-е изд., испр. и доп. - М.: ИНФРА-М, 2011. - 144 с.: (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=1#none)

3. Эконометрика: учебник / И. И. Елисеева. – М.: Проспект, 2010. – 288 с.

4. Уткин, В. Б. Эконометрика [Электронный ресурс] : Учебник / В. Б. Уткин; Под ред. проф. В. Б. Уткина. - 2-е изд. - М.: Издательско-торговая корпорация «Дашков и К°», 2012. - 564 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none)

5. Валентинов, В. А. Эконометрика [Электронный ресурс]: Практикум / В. А. Валентинов. - 3-е изд. - М.: Дашков и К, 2010. - 436 с.

(http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=3#none)

6. Эконометрика. Практикум: [Электронный ресурс] Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов.знание, 2014. - 329 с. (http://znanium.com/catalog.php?item=booksearch&code=%D1%8D%D0%BA%D0

%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%BA%D0%B0&page=4#none)

7.Тихомиров Н. П. Эконометрика: учебник. - М.: Экзамен, серия «Учебник Плехановской академии», 2007, -512 с.

8. Эконометрика: учебник / под ред. В. С. Мхитаряна. - М.: Проспект, 2008. -384 с.

9. Электронный курс “Econometrics and Public Policy (Advanced)”, Princeton University, URL: https://blackboard.princeton.edu/webapps/portal/frameset.jsp?tab_group=courses&url=%2Fwebapps%2Fblackboard%2Fexecute%2FcourseMain%3Fcourse_id%3D_214206_1

10. Электронный курс “Time Series Econometrics”, Princeton University, URL: http://sims.princeton.edu/yftp/Times05/;https://blackboard.princeton.edu/webapps/portal/frameset.jsp?tab_group=courses&url=%2Fwebapps%2Fblackboard%2Fexecute%2FcourseMain%3Fcourse_id%3D_52968_1.

Вопросы и задания для зачета

1. Каковы типы моделей и переменных, применяемых в эконометрике?
2. Что понимается под спецификацией модели?
3. Что понимается под верификацией модели?
4. Что такое функция регрессии?
5. Чем регрессионная модель отличается от функции регрессии?
6. Каковы основные причины наличия в регрессионной модели случайного отклонения?
7. Как осуществляется спецификация модели?
8. В чем состоит различие между теоретическим и эмпирическим уравнениями регрессии?
9. В чем суть метода наименьших квадратов?
10. Каковы формулы расчета коэффициентов эмпирического парного линейного уравнения регрессии по МНК?

11. Каковы предпосылки МНК? Каковы последствия их выполнимости или невыполнимости?

12. Действительно ли оценки коэффициентов регрессии будут иметь нормальное распределение, если случайные отклонения распределены нормально?

13. Действительно ли в любой линейной регрессионной модели, построенной по МНК, сумма случайных отклонений равна нулю?

14. В чем заключается суть коэффициента детерминации?

15. В каких пределах изменяется коэффициент детерминации?

16. Действительно ли для парной линейной регрессии коэффициент корреляции превосходит коэффициент детерминации?

17. Как записывается баланс для сумм квадратов отклонений результативного признака?

18. Что происходит, когда общая СКО равна остаточной? В каком случае общая СКО равна факторной?

19. Что такое число степеней свободы? Чему равны числа степеней свободы для различных СКО в парной регрессии?

20. Как используется F-статистика в регрессионном анализе?

21. Как рассчитать критерий Стьюдента для коэффициента регрессии в линейной модели парной регрессии?

22. В чем состоит "грубое" правило анализа статистической значимости коэффициентов регрессии?

23. Какая связь между t_b- и F- статистиками в парной линейной регрессии?

24. Какие этапы включает схема определения интервальных оценок коэффициентов регрессии?

25. В чем суть предсказания индивидуальных значений зависимой переменной?

26. В каком месте доверительный интервал прогноза по парной модели является наименьшим?Как записывается эмпирическое уравнение линейной модели множественной регрессии?

27. Что измеряют коэффициенты регрессии линейной модели множественной регрессии?

28. Какие требования предъявляются к факторам длявключения их в модель множественной регрессии?

29. Какой смысл приобретает сумма коэффициентов регрессии в производственных функциях?

30. Как в линейной модели множественной регрессии, записанной в стандартизованном виде, сравнить факторы по силе их воздействия на результат?

31. Как связаны стандартизованные коэффициенты регрессии с натуральными?

32. Как определяется статистическая значимость коэффициентов регрессии в линейной модели множественной регрессии?

33. Как строятся доверительные интервалы для параметров линейной модели множественной регрессии?

34. В чем недостаток использования коэффициента детерминации при оценке общего качества линейной модели множественной регрессии?

35. Как корректируется коэффициент детерминации?

36. Каково назначение частной корреляции при построении модели множественной регрессии?

37. Как определяется индекс множественной корреляции и какой он имеет смысл?

38. Каковы способы отбора факторов для включения в линейную модель множественной регрессии?

39. Как проверить обоснованность исключения части переменных из уравнения регрессии?

40. Как проверить обоснованность включения группы новых переменных в уравнение регрессии?

41. В чем различие терминов "коллинеарность" и "мультиколлинеарность"?

42. Что такое полная и частичная мультиколлинеарность?

43. Каковы причины и последствия мультиколлинеарности?

44. Как можно обнаружить мультиколлинеарность?

45. Каковы основные методы устранения мультиколлинеарности?

46. Какие значения парных коэффициентов корреляции свидетельствуют о наличии тесной связи независимых переменных?

47. Какой смысл имеет частный коэффициент корреляции?

48. В чем суть гомоскедастичности и гетероскедастичности? Каковы последствия гетероскедастичности?

49. Действительно ли, вследствие гетероскедастичности оценки перестают быть эффективными и состоятельными?

50. Какие критерии могут быть использованы для проверки гипотезы о гомоскедастичности регрессионных остатков?

51. В чем заключается тест Спирмена?

52. Какова схема теста Голдфелда-Квандта?

53. Каково предположение теста Парка?

54. В чем суть метода взвешенных наименьшихквадратов?

55. Какие типы преобразований применяются для устранения гетероскедастичности?

56. Что такое автокорреляция случайных отклонений?

57. Каковы основные причины и последствия автокорреляции?

58. Каковы основные методы обнаружения автокорреляции?

59. Что такое автокорреляционная функция?

60. В чем отличие положительной и отрицательной автокорреляции?

61. Какова основная идея метода рядов при обнаружении автокорреляции?

62. Как проводится тест Дарбина-Уотсона?

63. Как можно найти оценки регрессионных коэффициентов в случае линейной модели с коррелированными остатками?

64. В чем состоит авторегрессионная схема 1-го порядка?

65. В чем смысл поправки Прайса-Уинстена?

66. Какие статистические данные называют неоднородными?

67. Когда применяются фиктивные переменные?

68. В чем преимущества фиктивных переменных?

69. Как фиктивные переменные включаются в модель регрессии?

70. В чем суть ANOVA-моделей?

71. В чем суть ANCOVA-моделей?

72. В чем состоит правило применения фиктивных переменных?

73. Какой смысл имеет дифференциальный свободный член?

74. Какой смысл имеет дифференциальный угловой коэффициент?

75. В чем особенность моделей с переменной структурой?

76. Какова идея теста Чоу?

77. Как определяются коэффициенты эластичности по разным видам регрессионных моделей?

78. Какие показатели корреляции используются при нелинейных соотношениях рассматриваемых признаков?

79. В чем смысл средней ошибки аппроксимации и как она определяется?

80. Какие этапы содержит процедура построения тренд-сезонных моделей временных рядов?

81. В чем отличие аддитивной и мультипликативной моделей временных рядов?

82. Чему равна сумма сезонных компонент в аддитивной модели временного ряда?

83. Как осуществляется прогнозирование на основе трендовой и тренд-сезонной моделей временных рядов?

84. Как определяется модель ARMA?

85. Как интерпретируют параметры моделей авторегрессии?

86. Что означает стационарность временного ряда?

87. Какой стационарный процесс называется «белым шумом»?

88. Какие типы включают модели стационарных временных рядов?

89. Какие типы включают модели нестационарных временных рядов?

90. Как определяется ARIMA-модель?

91. Какие переменные называют предопределенными?

92. В чем отличие системы взаимозависимых уравнений от системы независимых уравнений?

93. В чем особенность системы рекурсивных уравнений?

94. Что такое структурная форма модели?

95. Что такое приведенная форма модели?

96. Почему нужна приведенная форма модели?

97. Что называют идентификацией модели?

98. В чем суть косвенного МНК?

99. Всегда ли можно применить косвенный МНК?

100. В чем суть двухшагового МНК и когда он применяется?

Задача 1. При исследовании корреляционной зависимости между ценой на нефть X и индексом нефтяных компаний Y получены следующие данные: $\bar{x} = 16{,}2$; $\bar{y} = 4000$; $\sigma_x^2 = 4$; $\operatorname{cov}(x, y) = 40$.

Задание:

1. По МНК оцените коэффициенты уравнений регрессии Y на X и X на Y.

2. Оцените коэффициент корреляции r_{yx} и коэффициент детерминации R^2.

Задача 2. Имеется следующая модель регрессии, характеризующая зависимость y от x:

$y = 8 - 7 \cdot x + e$. Известно, что $r_{yx}=-0{,}5$; $n=20$.

Задание:

1. Постройте доверительный интервал для коэффициента регрессии в этой модели: а) с вероятностью 90%, б) с вероятностью 99%.

2. Проанализируйте полученные результаты и поясните причины их различий.

Задача 3. По совокупности 30 торговых фирм изучается зависимость между ценой на товар, тыс. руб. (*X*) и прибылью, млн. руб. (*Y*). При оценке регрессионной модели были получены следующие промежуточные результаты:

$$\sum(y_i - \hat{y}_x)^2 = 39000; \sum(y_i - \bar{y})^2 = 120000$$

Задание:

1. Определите коэффициент детерминации.
2. Постройте таблицу дисперсионного анализа для расчета значения F-критерия Фишера.
3. Сравните фактическое значение F-критерия с табличным. Сделайте выводы.

Задача 4. Зависимость объема продаж, млн. руб. (*Y*) от расходов на рекламу, тыс. руб. (*X*) характеризуется по 12 предприятиям концерна следующим образом: $y=12{,}5+0{,}8*x;\ \sigma_x=5{,}4;\ \sigma_y=3{,}4.$

Задание:

1. Определите коэффициент корреляции.
2. Постройте таблицу дисперсионного анализа для оценки значимости уравнения в целом.
3. Найдите стандартную ошибку оценки коэффициента регрессии.
4. Оцените значимость коэффициента регрессии через t-критерий Стьюдента.
5. Определите доверительный интервал для коэффициента регрессии с вероятностью 0,95 и сделайте экономический вывод.

Задача 5. Уравнение регрессии, построенное по 15 наблюдениям, имеет вид:

$$\begin{array}{lcccc} y = & 10 - & 0{,}34x_1 + & 7{,}4x_2 - & ?\,x_3 \\ m_в & (9) & (\ ) & (2{,}2) & (4{,}5) \\ t_в & (\ ) & (-3{,}4) & (\ ) & (-4{,}1) \end{array}$$

Задание: Восстановите пропущенные значения и постройте доверительный интервал для b_3 с вероятностью 0,95.

Задача 6. По 30 заводам, выпускающим продукцию А, изучается зависимость потребления электроэнергии y (тыс. кВт*ч) от производства продукции – x_1(тыс. ед.) и уровня механизации труда – x_2(%). Данные приведены в таблице:

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	1000	27	r_{yx1}=0,77
x1	420	45	r_{yx2}=0,43
x2	41,5	18	r_{x1x2}=0,38

Задание:

1. Постройте уравнение множественной регрессии в стандартизованной и натуральной форме.

2. Определите показатели частной и множественной корреляции.

3. Найдите частные коэффициенты эластичности и сравните их с β-коэффициентами.

Задача 7. Уравнение регрессии в стандартизованных переменных выглядит так: $\hat{t}_y = -0{,}82 t_{x_1} + 0{,}65 t_{x_2} - 0{,}43 t_{x3}$.

При этом вариации всех переменных равны следующим величинам:

$$V_y = 32\%; V_{x1} = 38\%; V_{x2} = 43\%; V_{x3} = 35\%.$$

Задание: Сравнить факторы по степени влияния на результирующий признак и определить значения частных коэффициентов эластичности.

Задача 8. На основе помесячных данных за последние 4 года была построена аддитивная модель временного потребления тепла. Скорректированные значения сезонной компоненты приведены в таблице:

Январь	+ 30	май	- 20	сентябрь	- 10
февраль	+ 25	июнь	- 34	октябрь	?
март	+ 15	июль	- 42	ноябрь	+22
апрель	- 2	август	- 18	декабрь	+27

Уравнение тренда выглядит так:

$T = 350 + 1{,}3 \cdot t$

Задание: Определите значение сезонной компоненты за октябрь, а также точечный прогноз потребления тепла на 1 квартал следующего года.

Задача 9. На основе поквартальных данных за 16 лет построена мультипликативная модель некоторого временного ряда. Скорректированные значения сезонной компоненты равны:

I квартал – 1,4.

II квартал – 0,6.

III квартал – 0,5.

IV квартал - ?

Уравнение тренда имеет вид:

$T = 10{,}4 - 0{,}2 \cdot t$

Задание: Определите значение сезонной компоненты за IV квартал и прогноз на II и III кварталы следующего года.

Задача 10. На основе квартальных данных объемов продаж 1999 – 2004 гг. была построена аддитивная модель временного ряда. Трендовая компонента имеет вид: $T = 260 + 3 \cdot t$

Показатели за 2004 г. приведены в таблице:

Квартал	Фактический объем продаж	Компонента аддитивной модели		
		трендовая	сезонная	случайная
1	270	T_1	S_1	-9
2	Y_2	T_2	10	+4
3	310	T_3	40	E_3
4	Y_4	T_4	S_4	E_4
ИТОГО:	1200			

Задание: Заполните недостающие данные в таблице.

Задача 11. Имеются следующие данные об уровне безработицы y_t (%) за 8 месяцев:

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

Задание:

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

2. Обоснуйте выбор уравнивания тренда и определите его параметры.

3. Интерпретируйте полученные результаты.

Задача 12. В таблице указаны остатки регрессии:

Год	Остатки	Год	Остатки	Год	Остатки
1996	- 0,6	2000	0,0	2004	0,0
1997	0,0	2001	0,2	2005	0,2
1998	- 0,1	2002	- 0,1	2006	0,2
1999	0,8	2003	- 0,1	2007	- 0,1

Задание: Применив критерий Дарбина-Уотсона, сделайте вывод об отсутствии или наличии автокорреляции.

Задача 13. Модель Менгеса имеет следующий вид:

$Y_t = a_1 + b_{11}Y_{t-1} + b_{12}I_t + e_1,$

$I_t = a_2 + b_{21}Y_t + b_{22}Q_t + e_2,$

$C_t = a_3 + b_{31}Y + b_{32}C_{t-1} + b_{33}P_t + e_3,$

$Q_t = a_4 + b_{41}Q_{t-1} + b_{42}R_t + e_4.$

где Y – национальный доход;

C – расходы на личное потребление;

I – чистые инвестиции;

Q – валовая прибыль экономики;

P – индекс стоимости жизни;

R – объем продукции промышленности;

t – текущий период;

t – 1 – предыдущий период.

Задание: Применив необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений модели. Определите метод оценки параметров модели. Запишите приведенную форму модели.

Задача 14. Одна из версий модели Кейнса имеет вид:

$$C_t = a_1 + b_{11}Y_t + b_{12}Y_{t-1} + \varepsilon_1,$$
$$I_t = a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2,$$
$$Y_t = C_t + I_t + G_t,$$

где C - расходы на потребление;

Y - доход;

I - инвестиции;

G - государственные расходы;

t - текущий период;

$t-1$ - предыдущий период.

Задание: Применив необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений модели. Определите метод оценки параметров модели. Запишите приведенную форму модели.

Задача 15. Имеется следующая структурная модель:

$Y_1 = b_{12}Y_2 + a_{11}X_1 + a_{12}X_2,$

$Y_2 = b_{21}Y_1 + b_{23}Y_3 + a_{22}X_2,$

$Y_3 = b_{32}Y_2 + a_{31}X_1 + a_{33}X_3.$

Приведенная форма исходной модели имеет вид:

$Y_1 = 2X_1 – 4X_2 + 10X_3,$

$Y_2 = X_1 + 3X_2 + 2X_3,$

$Y_3 = -6X_1 + 7X_2 + 6X_3.$

Задание:

1. Проверьте структурную форму модели на идентификацию.

2. Определите структурные коэффициенты первого уравнения модели.