КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

ИНСТИТУТ ЭКОНОМИКИ И ФИНАНСОВ

Е. И. КАДОЧНИКОВА

ЭКОНОМЕТРИКА

Конспект лекций

Казань – 2013

Кадочникова Е. И.

Эконометрика. Конспект лекций / Е. И. Кадочникова; Каз. федер. ун-т. – Казань, 2013. – 106 с.

В предлагаемых лекциях изучаются теоретические и методологические вопросы эконометрического моделирования с позиции четырех основных разделов: линейная модель регрессии и метод наименьших квадратов, обобщенный метод наименьших квадратов, модели временных рядов, системы одновременных уравнений. Приведены основные положения по верификации эконометрических моделей, в частности, проверке соблюдения предпосылок метода наименьших квадратов. Рассмотрены основные приемы тестирования остатков регрессии на гетероскедастичность и автокорреляцию, описаны особенности построения моделей временных рядов, вопросы проверки временных рядов на стационарность. Отдельное внимание уделено вопросам построения и оценивания систем одновременных уравнений.

Для этого курса имеется электронная версия –

http://tulpar.kpfu.ru/course/view.php?id=382

Принято на заседании кафедры статистики, эконометрики и естествознания

Протокол № 3 от 28.11.2013

Содержание

•Лекция 1(1). Эконометрика как научная дисциплина

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 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.

Вопросы для изучения:

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

Цели, предмет, задачи эконометрики

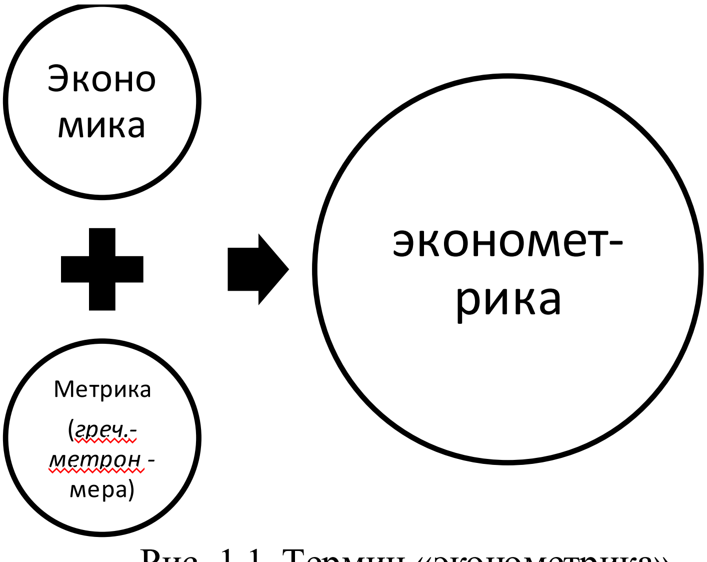

Рис. 1.1. Термин «эконометрика»

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

«Эконометрика – это наука, которая на базе экономической теории, экономической статистики, экономических измерений и математикостатистического инструментария придает количественное выражение качественным закономерностям, обусловленным экономической теорией» (С. А. Айвазян).

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

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

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики.

«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).

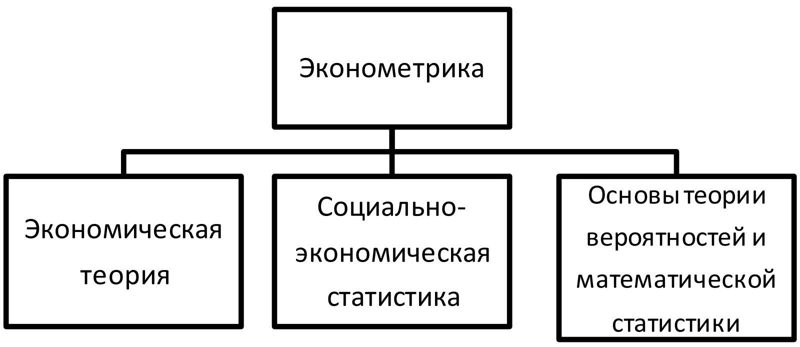

Рис. 1.2. Три составляющие эконометрики

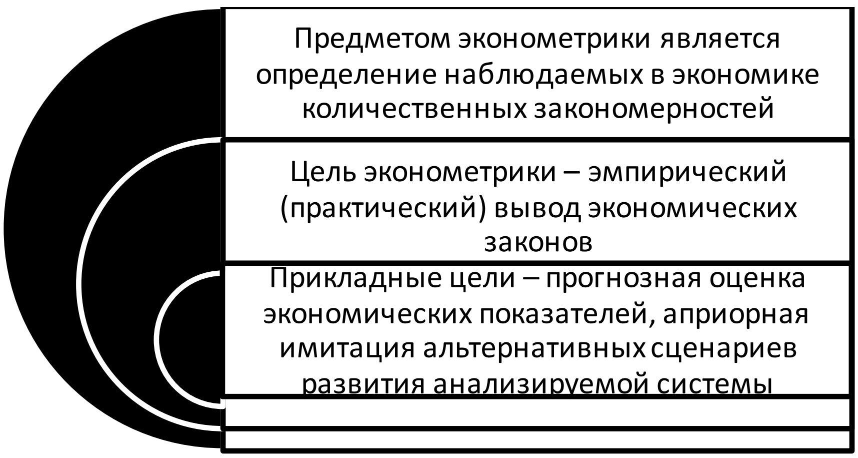

Рис. 1.3. Предмет и цели эконометрики

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

Разделы:

- линейная модель регрессии и МНК;
- обобщенная линейная модель регрессии и ОМНК;
- статистический анализ временных рядов;
- анализ систем одновременных уравнений.

Особенности эконометрического метода:

- 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.

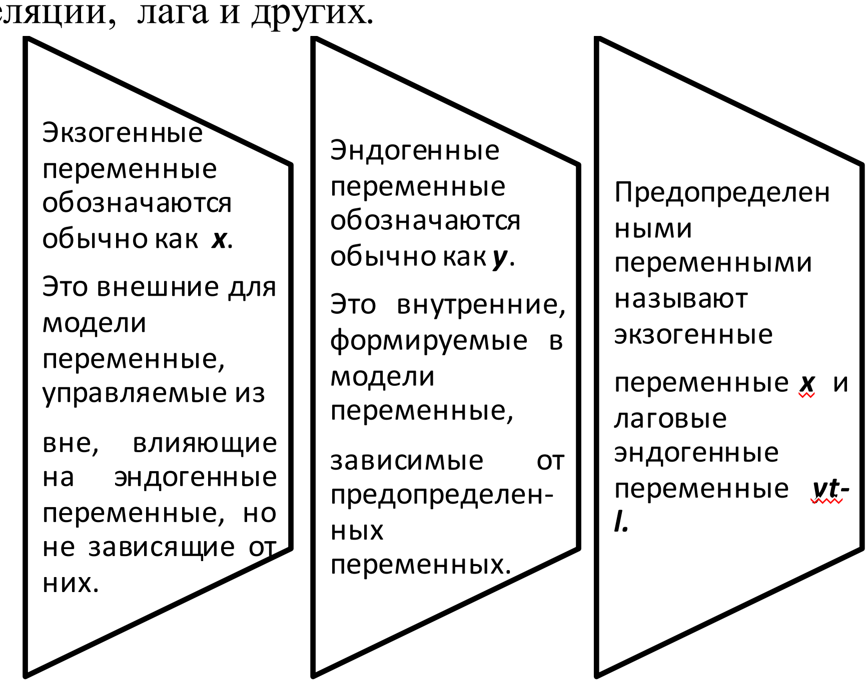

Рис. 1.4. Характеристика переменных

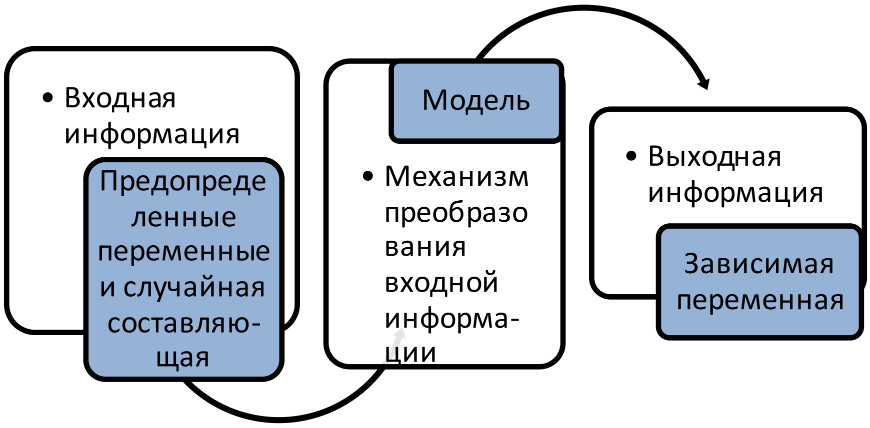

Рис. 1.5. Взаимодействие переменных

Модели временных рядов:

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

Теоретическая функция регрессии (общий вид):

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

Теоретическая линейная модель множественной регрессии:

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

Система одновременных уравнений:

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

Типы данных в эконометрике:

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

Выделяют следующие этапы: постановочный; априорный; спецификация модели; информационный; идентификация модели; верификация модели; интерпретация результатов.

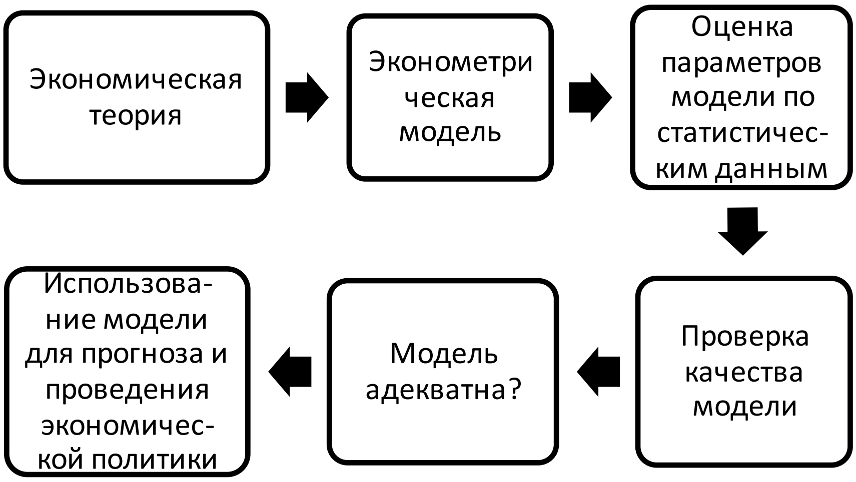

Рис. 1.6. Последовательность эконометрического моделирования

Вопросы для самоконтроля

1. Что измеряет эконометрика?
2. Назовите основные цели эконометрики.
3. В чем состоят предмет и задачи эконометрики?
4. Какие типы моделей и переменных применяют 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?
11. Приведите примеры случайных событий в экономике. Можно ли дать им вероятностное описание?
12. Перечислите основные свойства математического ожидания.
13. Перечислите основные свойства дисперсии.
14. Дайте определение ковариации. Как определяется коррелированность и некоррелированность случайных величин?
15. Что такое генеральная совокупность и выборка?
16. Как вычисляются основные числовые характеристики по результатам выборки: выборочные среднее, дисперсия, среднее квадратическое отклонение?
17. Что такое функция распределения случайной величины? Приведите ее свойства.

18. Что такое плотность вероятности случайной величины? Приведите ее свойства.

Задания для практики

Задача 1. В лотерее разыгрывается: автомобиль стоимостью 5000 ден. ед., 4 телевизора стоимостью 250 ден. ед., 5 видеомагнитофонов стоимостью 200 ден. ед. Всего продается 1000 билетов по 7 ден. ед.

Задание:

1)составить закон распределения чистого выигрыша, полученного участником лотереи;

2) вычислить математическое ожидание для случайной величины – чистого выигрыша;

3) вычислить дисперсию и среднеквадратическое отклонение случайной величины.

Задача 2. Пусть X, Y – годовые дивиденды от вложений в отрасли А и В соответственно. Риск от вложений характеризуется дисперсиями: D(X) = 16, D(Y) = 9. Коэффициент корреляции ρ(X,Y) = -0,6.

Задание: что менее рискованно, вкладывать деньги в обе отрасли в соотношении 30% на 70% или только в отрасль В?

Задача 3. Пусть X, Y - годовые дивиденды от вложений денежных средств в акции компаний А и В соответственно. Риск от вложений характеризуется дисперсиями $D(X)=25$, $D(Y)=16$. Коэффициент корреляции ρ_{xy} =0,8.

Задание: что менее рискованно, вкладывать деньги в обе компании в соотношении 25% и 75% или только в компанию В?

Задача 4. Прибыль в отрасли имеет нормальный закон распределения со средним значением 1млн$ и средним квадратическим отклонением 0,25 млн. $.

Задание: что вероятнее, получить прибыль не более чем 0,8 млн. $ или в пределах от 1,2 млн. $ до 1,5 млн. $?

Глоссарий

Предопределенные переменные – это экзогенные переменные и лаговые эндогенные переменные.

Цель эконометрики – эмпирический (практический) вывод экономических законов.

Экзогенные переменные - это внешние для модели переменные, управляемые из вне, влияющие на эндогенные переменные, но не зависящие от них.

Эконометрика – это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов, которые раскрыты и обоснованы экономической теорией.

Эндогенные переменные - это внутренние, формируемые в модели переменные, зависимые от предопределенных переменных.

Использованные информационные ресурсы

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

2. Бородич С. А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006.- Гл. 1,2,3.

3. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 1.

Лекция 1(2)

•Лекция 1(2). Линейная модель парной регрессии и метод наименьших квадратов

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 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.

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.

Спецификация линейной модели парной регрессии

Основная цель регрессионного анализа – оценка функциональной зависимости между независимыми переменными X и условным математическим ожиданием зависимой переменной Y.

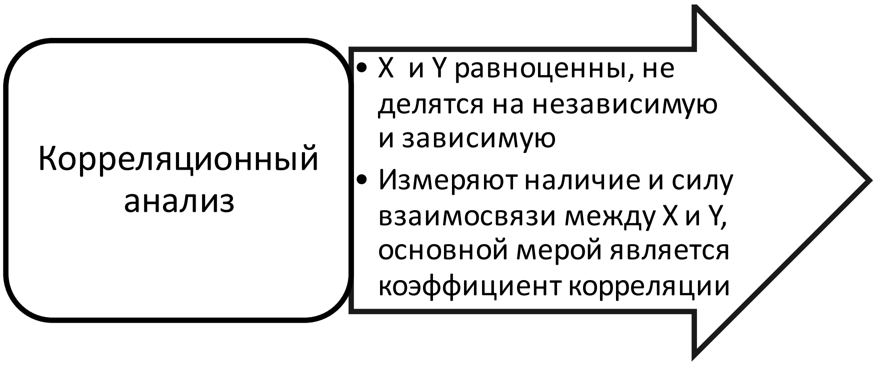

Рис. 2.1. Суть регрессионного анализа

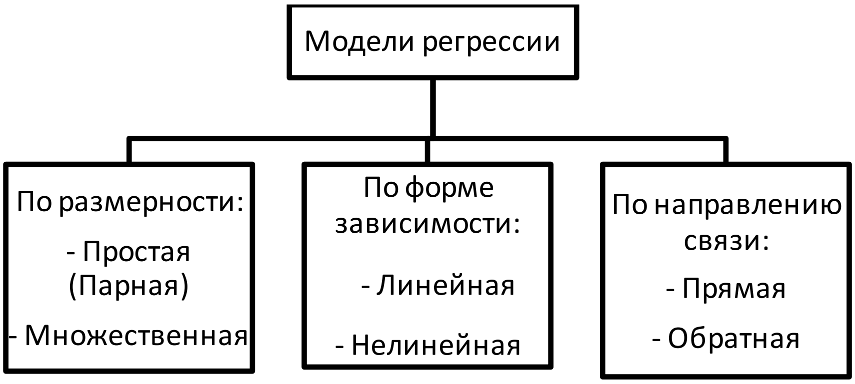

Рис. 2.2. Виды регрессии

Простая (парная) регрессия представляет собой модель, где теоретическое (среднее) значение зависимой переменной Y рассматривается как функция одной независимой переменной X: $Y_x = f(x)$

Множественная регрессия представляет собой модель, где теоретическое (среднее) значение зависимой переменной Y рассматривается как функция нескольких независимых переменных X_1, X_2,...X_m: $Y_x = f(x_1, x_2, ..., x_m)$

Спецификация модели - формулирование вида модели, исходя из соответствующей теории связи между переменными. Определяется состав переменных и математическая функция для отражения связи между ними.

Спецификация линейной модели (уравнения) парной регрессии: $Y_i = Y_{xi} + \varepsilon_i$

где Y_i - фактическое значение зависимой переменной Y;

Y_{xi} - теоретическое (среднее) значение зависимой переменной Y;

ε_i - случайная величина (остаток регрессии).

Теоретическое уравнение регрессии (гипотетически для генеральной совокупности): $Y_i = \alpha + \beta \cdot x_i + \varepsilon_i$

где α – свободный коэффициент;

β - коэффициент регрессии;

ε_i – случайное отклонение (возмущение).

Случайное отклонение включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Источники его присутствия в модели: спецификация модели, выборочный характер исходных данных, особенности измерения переменных.

Эмпирическое уравнение регрессии (для выборки наблюдений): $Y_i = a + b \cdot x_i + e_i$

где a – эмпирическая (выборочная) оценка свободного коэффициента;

b - эмпирическая (выборочная) оценка коэффициента регрессии;

e_i – эмпирическая (выборочная) оценка теоретического случайного отклонения ε (остаток регрессии).

Метод наименьших квадратов (МНК)

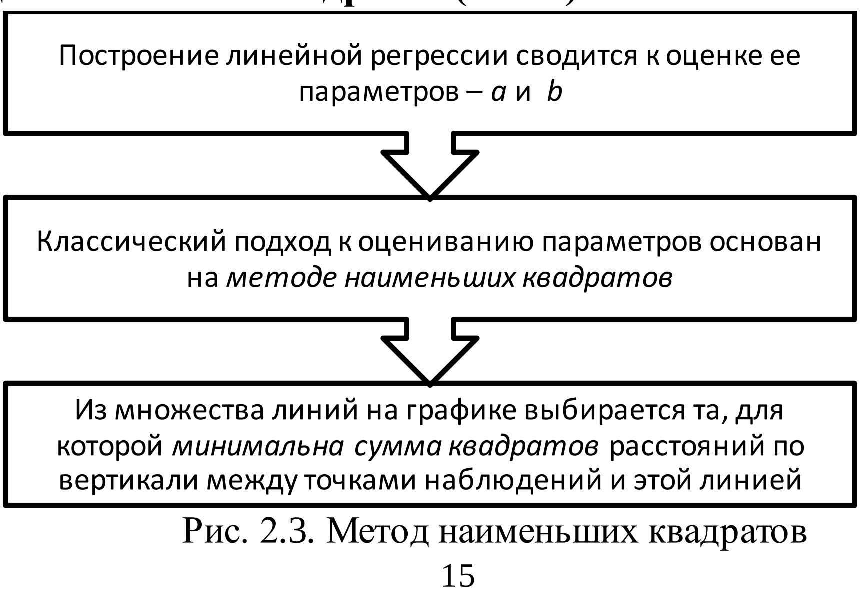

Рис. 2.3. Метод наименьших квадратов

Суть метода наименьших квадратов (МНК) - оценки параметров таковы, что сумма квадратов отклонений фактических значений зависимой переменной Y_i от расчетных (теоретических) Y_x минимальна:

$$\sum_{i=1}^{n}(y_i - y_{xi})^2 \to \min$$

Оценка параметров регрессии:

$$S = \sum (y_i - y_{x\,i})^2 = \sum (y - a - b \cdot x)^2;$$

$$\frac{dS}{da} = -2\sum y + 2 \cdot n \cdot a + 2 \cdot b \sum x = 0;$$

$$\frac{dS}{db} = -2\sum y \cdot x + 2 \cdot a \sum x + 2 \cdot b \sum x^2 = 0.$$

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

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

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

Предпосылки МНК и свойства МНК-оценок

В силу несовпадения статистической базы для генеральной совокупности и выборки оценки параметров регрессии a и b отличаются от теоретических коэффициентов α и β и не позволяют сделать вывод, насколько точно эмпирическое уравнение регрессии соответствует уравнению для всей генеральной совокупности.

Доказано, что надежность оценок параметров регрессии существенно зависит от свойств случайного отклонения ε.

Для получения наилучших МНК-оценок необходимо, чтобы выполнялся ряд предпосылок относительно ε.

Предпосылки МНК:

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

2. Дисперсия случайных отклонений εi постоянна. Выполнение предпосылки называется гомоскедатичностью, нарушение – гетероскедастичностью: $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$

3. Случайные отклонения ε_i и ε_j являются независимыми друг от друга для $i \neq j$. Выполнение данной предпосылки говорит об отсутствии автокорреляции, нарушение – о присутствии автокорреляции:

$$\sigma_{\varepsilon_i \varepsilon_j} = \operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

$$\sigma_{\varepsilon_i \varepsilon_j} = \operatorname{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2, i = j$$

4. Случайное отклонение должно быть независимо от объясняющих переменных: $\sigma_{\varepsilon_i x_i} = \text{cov}(\varepsilon_i, x_i) = 0$

5. Модель линейна относительно параметров.

Если предпосылки МНК выполнены, то МНК-оценки регрессии обладают следующими свойствами:

1.Оценки являются несмещенными: $M(a) = \alpha, M(b) = \beta$

2. Оценки состоятельны, так как их дисперсия при увеличении выборки стремится к нулю: $D(a) \to 0, D(b) \to 0, n \to \infty$

3.Оценки эффективны, имеют наименьшую дисперсию по сравнению с другими оценками, линейными относительно зависимой переменной: $D(a) \to D_{\min}, D(b) \to D_{\min}$

Вопросы для самоконтроля

1. Что такое функция регрессии?
2. Чем регрессионная модель отличается от функции регрессии?
3. Назовите 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. Приведите формулы расчета коэффициентов эмпирического парного линейного уравнения регрессии по МНК.
8. Перечислите 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?

Задания для практики

Задача 1. При приеме на работу семи кандидатам было предложено два теста. Результаты тестирования в баллах приведены в таблице:

Тест	Результаты тестирования кандидатов						
	1-й	2-й	3-й	4-й	5-й	6-й	7-й
1	31	82	25	26	53	30	29

2	21	55	8	27	32	42	26

Задание: Найти коэффициент ранговой корреляции Спирмена между результатами тестирования по двум тестам.

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

Задача 3. По 12 регионам России приводятся данные о среднедушевом прожиточном минимуме в день одного трудоспособного x (руб.) и среднедневной заработной плате y (руб.):

Номер региона	1	2	3	4	5	6	7	8	9	10	11	12
x, руб	78	82	87	79	89	106	67	88	73	87	76	115
y, руб	133	148	134	154	162	195	139	158	152	162	159	173

Задание: построить поле корреляции и сформулировать предложение о форме связи переменных x и y, построить уравнение линейной парной регрессии.

Задача 4. Имеются данные о финансовых результатах по территориям Приволжского федерального округа:

Территории Приволжского федерального округа	Сальдированный финансовый результат (прибыль) за год, млрд. руб.	Инвестиции в основной капитал за год, млрд. руб.
Республика Башкортостан	43,4	62,4
Республика Марий Эл	0,6	5,8
Республика Мордовия	1,6	10,4
Республика Татарстан	70,0	86,6
Республика Удмуртия	6,4	15,4
Чувашская республика	3,0	14,2
Кировская обл.	3,2	9,5
Нижегородская обл.	24,2	48,5
Оренбургская обл.	19,8	27,7
Пензенская обл.	1,8	10,7
Пермская обл.	43,5	48,2
Самарская обл.	2,8	55,0
Саратовская обл.	8,3	23,8
Ульяновская обл.	1,4	11,3

Задание: для изучения зависимости прибыли от размера инвестиций в основной капитал рассчитайте параметры a и b парной линейной функции $y_x = a + b \cdot x$. По уравнению регрессии рассчитайте теоретические значения результата ($\hat{y}_x$), по ним постройте теоретическую линию регрессии, определите среднюю ошибку аппроксимации и оцените ее величину.

Глоссарий

Множественная регрессия представляет собой модель, где теоретическое (среднее) значение зависимой переменной Y рассматривается как функция нескольких независимых переменных $X_1, X_2, \ldots X_m$.

Параметризация модели – выражение в математической форме взаимосвязи между переменными модели, формулирование исходных предпосылок и ограничений модели.

Парная регрессия представляет собой модель, где теоретическое (среднее) значение зависимой переменной Y рассматривается как функция одной независимой переменной X.

Спецификация модели - формулирование вида модели, исходя из соответствующей теории связи между переменными.

Цель регрессионного анализа – оценка функциональной зависимости между независимыми переменными X и условным математическим ожиданием зависимой переменной Y.

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11787
2. Бородич С. А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006.- Гл. 4, 5.
3. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 2.

Лекция 1(3)

Лекция 1(3). Экономическая и статистическая интерпретация линейной модели парной 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.

Вопросы для изучения:

1. Экономическая интерпретация параметров модели.
2. Коэффициенты корреляции и детерминации в линейной модели парной регрессии.
3. Проверка качества модели линейной парной регрессии (верификация модели).
4. Интервалы прогноза по линейному уравнению регрессии.

Экономическая интерпретация параметров модели

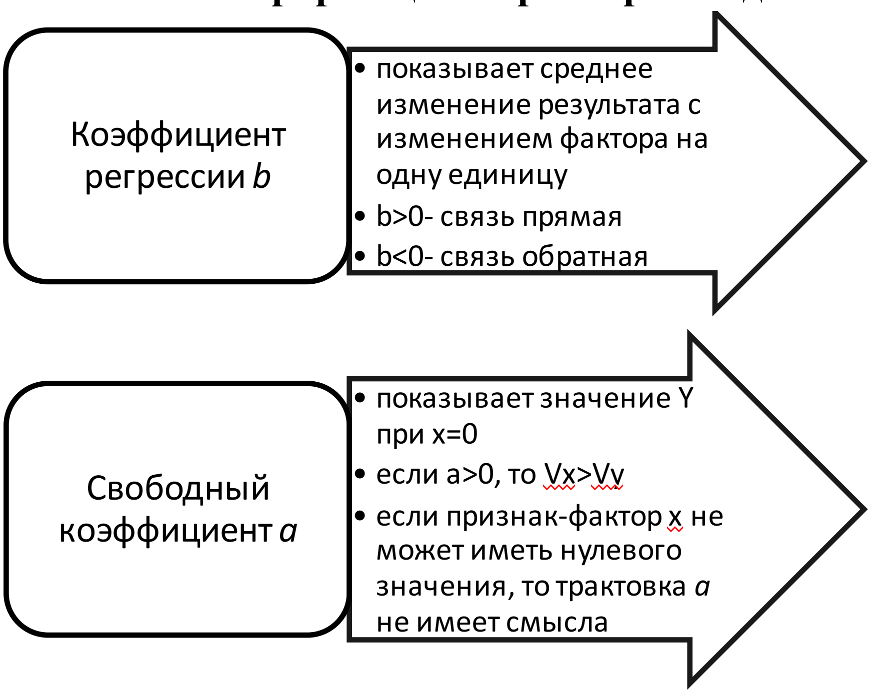

Рис. 3.1. Интерпретация параметров модели

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.

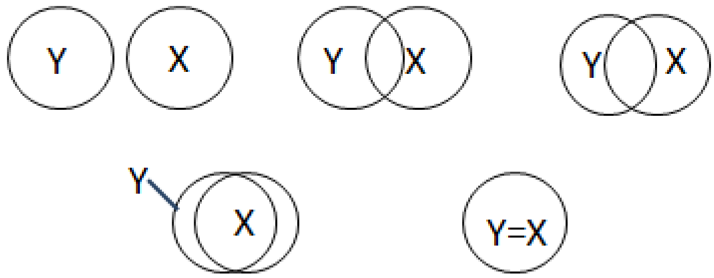

Рис. 3.2. Диаграмма Венна

Линейный коэффициент парной корреляции:

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

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

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): $\sum (y_i - \overline{y})^2$
- регрессионная (ESS): $\sum (y_x - \overline{y})^2$
- остаточная (RSS): $\sum (y_i - y_x)^2$

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

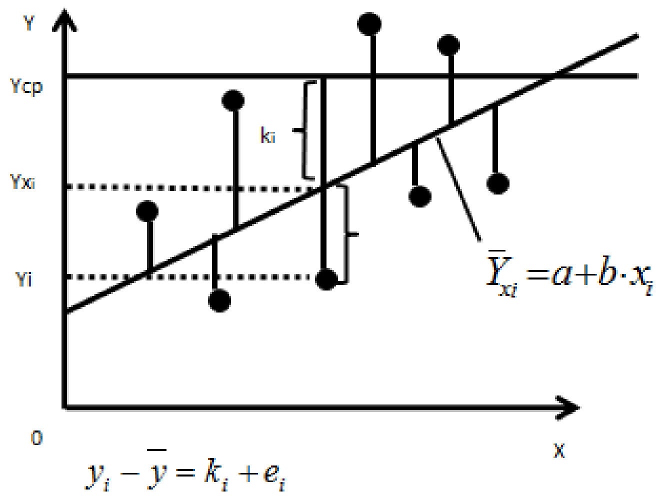

Рис. 3.3. Геометрическая интерпретация

Выборочные оценки дисперсий:

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

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

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

Коэффициент детерминации:

$$R^2 = \frac{\sum (y_x - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (y_i - y_x)^2}{\sum (y_i - \bar{y})^2}$$

$$R^2 = r^2_{yx}; 0 \le R^2 \le 1$$

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

Проверка качества модели линейной парной регрессии (верификация модели)

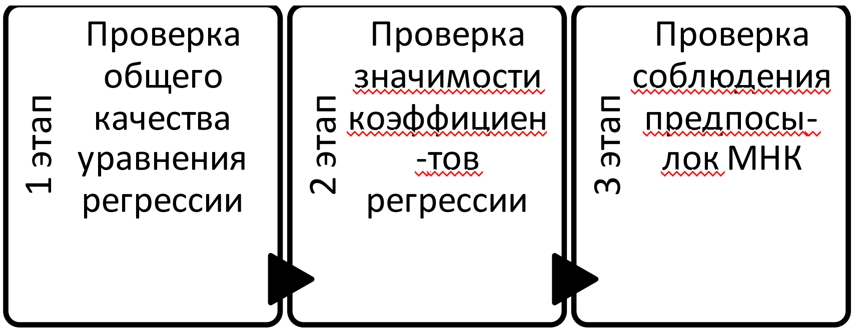

Рис. 3.4. Этапы проверки качества модели

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.

$$H_0 : \beta = 0$$

$$H_1 : \beta \ne 0$$

$$|t_b| > t_{\alpha/2, n-2} \Rightarrow H_1$$

$$|t_b| < t_{\alpha/2, n-2} \Rightarrow H_0$$

$${}_b = \frac{b}{m_b}; t_a = \frac{a}{m_a}; t_r = \frac{r}{m_r} = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}$$

$$m_b = \sqrt{\frac{\sum (y-y_x)^2 /(n-2)}{\sum (x-\overline{x})^2}} = \sqrt{\frac{S^2{}_{RSS}}{\sum (x-\overline{x})^2}} = \frac{S_{RSS}}{\sigma_x \sqrt{n}}$$

$$m_a = \sqrt{\frac{\sum (y-y_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n\sum (x-\overline{x})^2}}$$

$$m_r = \sqrt{\frac{1-r^2}{n-2}}; t^2{}_r = t^2{}_b = F$$

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

По уравнению регрессии определяется прогнозное значение зависимой переменной Y_x ($\overline{Y}_{xp}$) путем подстановки в уравнение прогнозного значения $X_{пр}$. Точечный прогноз дополняется интервальной оценкой прогноза $\overline{Y}_{xp}$.

Предсказание среднего и индивидуального значения зависимой переменной:

$$\overline{Y}_{xp} - m_{Yxp} \le Y^* \le \overline{Y}_{xp} + m_{Yxp};$$

$$a + b \cdot x_{ïð} \pm t_{\alpha/2,n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\overline{x} - x_{ïð})^2}{\sum (x_i - \overline{x})^2}}$$

$$a + b \cdot x_{ïð} \pm t_{\alpha/2,n-2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(\overline{x} - x_{ïð})^2}{\sum (x_i - \overline{x})^2}}$$

где m_{Yxp} – стандартная ошибка точечного прогноза;

S^2 – остаточная дисперсия на одну степень свободы;

t – случайная величина, имеющая распределение Стьюдента с заданной вероятностью.

Вопросы для самоконтроля

1. Объясните суть коэффициента детерминации. В каких пределах он изменяется?
2. Опишите схему проверки гипотезы о величине коэффициента детерминации.
3. Как определяются стандартные ошибки коэффициентов регрессии?
4. Опишите схему проверки гипотез о величинах коэффициентов регрессии.
5. В чем суть статистической значимости коэффициентов регрессии?
6. Приведите схему определения интервальных оценок коэффициентов регрессии.
7. В чем суть предсказания значений зависимой переменной.
8. Объясните суть коэффициента детерминации. В каких пределах изменяется коэффициент детерминации.
9. Действительно ли для парной линейной регрессии коэффициент корреляции превосходит коэффициент детерминации?

Задания для практики

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: по МНК оцените коэффициенты уравнений регрессии Y на X и X на Y. 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: постройте доверительный интервал для коэффициента регрессии в этой модели: а) с вероятностью 90%, б) с вероятностью 99%. 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: определите коэффициент детерминации. Постройте таблицу дисперсионного анализа для расчета значения F-критерия Фишера. 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: определите коэффициент корреляции. Постройте таблицу дисперсионного анализа для оценки значимости уравнения в целом. Найдите стандартную ошибку оценки коэффициента регрессии. Оцените значимость коэффициента регрессии через t-критерий Стьюдента. Определите доверительный интервал для коэффициента регрессии с вероятностью 0,95 и сделайте экономический вывод.

Глоссарий

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

Идентификация модели – проведение статистического анализа модели и оценивания качества ее параметров.

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

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

Коэффициент детерминации – это показатель, который определяет долю разброса зависимой переменной Y, объясняемую регрессией Y на X.

Линейный коэффициент парной корреляции – это показатель тесноты статистической взаимосвязи между переменными Y и X.

Парный коэффициент регрессии показывает, на какую величину в среднем изменится результативный признак Y, если переменную X увеличить на единицу измерения.

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

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11797
http://tulpar.kpfu.ru/mod/resource/view.php?id=11798

2. Бородич С.А. Эконометрика: учебное пособие. - Мн.: Новое знание, 2006. – Гл. 4, 5.

3. Эконометрика: учеб. /под ред. И.И. Елисеевой. -М.: Проспект, 2010.- Гл. 2.

Лекция 2(1)

Лекция 2(1). Линейная модель множественной регрессии, оценка ее параметров

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.

Вопросы для изучения:

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

Линейная модель множественной регрессии. Эмпирическая форма записи.

Множественная регрессия представляет собой модель, где среднее значение зависимой переменной Y рассматривается как функция нескольких независимых переменных X_j: $Y = f(x_1, x_2, x_3, ... x_m) + \varepsilon$

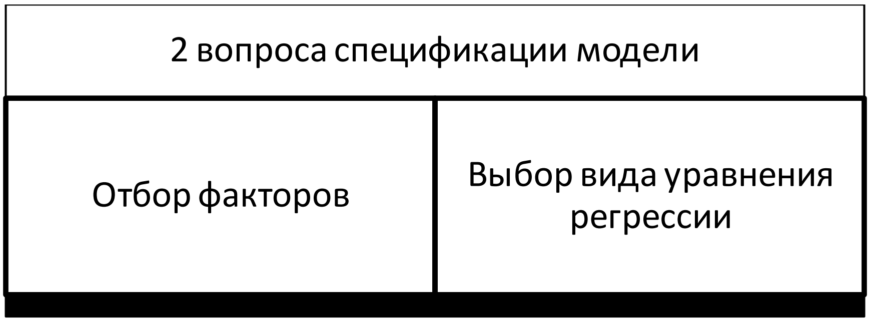

Рис. 4.1. Проблемы спецификации модели

Линейная модель множественной регрессии:

$$Y = a_0 + b_1 x_1 + b_2 x_2 + ... + b_m x_m + \varepsilon$$

Факторы, включаемые во множественную регрессию:

- должны быть количественно измеримы;
- не должны быть коррелированы между собой и тем более находиться в точной функциональной связи.

Нелинейные модели множественной регрессии:

$$Y = a_0 \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot ... x_m^{b_m} \cdot \varepsilon$$

$$Y = e^{a_0 + b_1 x_1 + ... + b_m x_m + \varepsilon}$$

$$Y = a_0 + \frac{b_1}{x_1} + \frac{b_2}{x_2} + ... + \frac{b_m}{x_m} + \varepsilon$$

$$Y = a_0 + b_1 x_1^2 + b_2 x_2^2 + ... + b_m x_m^2 + \varepsilon$$

Оценка параметров модели с помощью МНК

МНК-оценки множественной регрессии:

$$e_i = y_i - a_0 - b_1 x_1 - ... - b_m x_m$$

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (a_0 + \sum_{j=1}^{m} b_j x_{ij}))^2 \to \min$$

$$\begin{cases} \dfrac{\partial Q}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - (a_0 + \sum_{j=1}^{m} b_j x_{ij})), \\ \dfrac{\partial Q}{\partial b_j} = -2\sum_{i=1}^{n}(y_i - (a_0 + \sum_{j=1}^{m} b_j x_{ij})) \cdot x_{ij} \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n}(y_i - (a_0 + \sum_{j=1}^{m} b_j x_{ij})) = 0, \\ \sum_{i=1}^{n}(y_i - (a_0 + \sum_{j=1}^{m} b_j x_{ij})) \cdot x_{ij} = 0 \end{cases}$$

$$\begin{cases} \sum y = n \cdot a_0 + b_1 \cdot \sum x_1 + b_2 \cdot \sum x_2 + \ldots + b_m \cdot \sum x_m, \\ \sum y \cdot x_1 = a_0 \cdot \sum x_1 + b_1 \sum x_1^{\ 2} + b_2 \sum x_1 \cdot x_2 + \ldots + b_m \sum x_m x_1, \\ \ldots\ldots \\ \sum y \cdot x_m = a_0 \sum x_m + b_1 \sum x_1 \cdot x_m + b_2 \sum x_2 \cdot x_m + \ldots + b_m \sum x_m^{\ 2} \end{cases}$$

Стандартизованные коэффициенты регрессии:

$$t_y = \beta_1 \cdot t_{x1} + \beta_2 \cdot t_{x2} + \ldots + \beta_m \cdot t_{xm} + \varepsilon$$

$$t_y = \frac{y - \bar{y}}{\sigma_y}; t_{xj} = \frac{x_j - \bar{x}_j}{\sigma_{x_j}};$$

$$t_y = t_{x_j} = 0; \sigma_{ty} = \sigma_{tx} = 1$$

Благодаря тому, что в стандартизованном уравнении все переменные заданы как центрированные и нормированные, β-коэффициенты сравнимы между собой. Сравнивая друг с другом β-коэффициенты, можно ранжировать факторы по силе их влияния на зависимую переменную Y. Коэффициенты «чистой» регрессии (b_j) несравнимы между собой.

$$\begin{cases} Ryx_1 = \beta_1 + \beta_2 \cdot Rx_2x_1 + \beta_3 \cdot Rx_3x_1 + \ldots + \beta_m \cdot Rx_mx_1, \\ Ryx_2 = \beta_1 \cdot x_2x_1 + \beta_2 + \beta_3 \cdot Rx_3x_2 + \ldots + \beta_m \cdot Rx_mx_2, \\ \ldots\ldots \\ Ryx_m = \beta \cdot x_mx_1 + \beta_2 \cdot Rx_mx_2 + \beta_3 \cdot Rx_3x_m + \ldots + \beta_m \end{cases}$$

Стандартизованные коэффициенты регрессии показывают, на сколько сигм изменится в среднем результат, если соответствующий фактор x_j изменится на одну сигму при неизменном среднем уровне других факторов.

В парной зависимости стандартизованный коэффициент регрессии есть линейный коэффициент корреляции: $\beta = r_{xy}$

Во множественной регрессии зависимость следующая: $b_j = \beta_j \dfrac{\sigma_y}{\sigma_{x_j}}$

Частное уравнение регрессии связывает результативный признак Y с соответствующим фактором X_j при закреплении других факторов на среднем уровне и характеризует изолированное влияние фактора X_j на результат.

$$y_{x_1 \cdot x_2,\dots x_m} = a_0 + b_1 x_1 + b_2 \bar{x}_2 + \dots + b_m \bar{x}_m + \varepsilon;$$
$$y_{x_2 \cdot x_1, x_3 \dots x_m} = a_0 + b_1 \bar{x}_1 + b_2 x_2 + \dots + b_m \bar{x}_m + \varepsilon;$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots;$$
$$y_{x_m \cdot x_1,\dots x_{m-1}} = a_0 + b_1 \bar{x}_1 + b_2 \bar{x}_2 + \dots + b_m x_m + \varepsilon.$$

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

Частный коэффициент эластичности показывает, на сколько % изменяется в среднем результативный признак Y при изменении фактора Xj на 1% и при неизменных других факторах, включенных в модель.

$$Э_{y_{x_j}} = b_j \cdot \frac{x_j}{y_{x_j \cdot x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_m}}$$

Показатели качества множественной регрессии

Индекс множественной корреляции независимо от формы связи оценивает тесноту совместного влияния факторов на результативный признак Y:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\sum (y - y_{x_1x_2\dots x_m})^2}{\sum (y - \bar{y})^2}}$$

$$0 \le R_{yx_1x_2\dots x_m} \le 1$$

При линейной регрессии: $R_{yx_1x_2\dots x_m} = \sqrt{\sum \beta_{x_j} \cdot r_{yx_j}}$

Коэффициент детерминации:

$$R^2{}_{yx_1x_2\dots x_m} = 1 - \frac{\sum (y - y_{x_1x_2\dots x_m})^2}{\sum (y - \bar{y})^2}$$

Скорректированный коэффициент детерминации:

$$R^2{}_{yx_1x_2\dots x_m} = 1 - \frac{\sum (y - y_{x_1x_2\dots x_m})^2 / (n - m - 1)}{\sum (y - \bar{y})^2 / (n - 1)}$$

Когда m - число параметров при X_j – приближается к объему наблюдений (n), то остаточная дисперсия будет близка к нулю и R^2 приблизится к 1 даже при слабой связи факторов с результатом. Скорректированный R^2 содержит поправку на число степеней свободы, что не допускает возможного преувеличения тесноты связи.

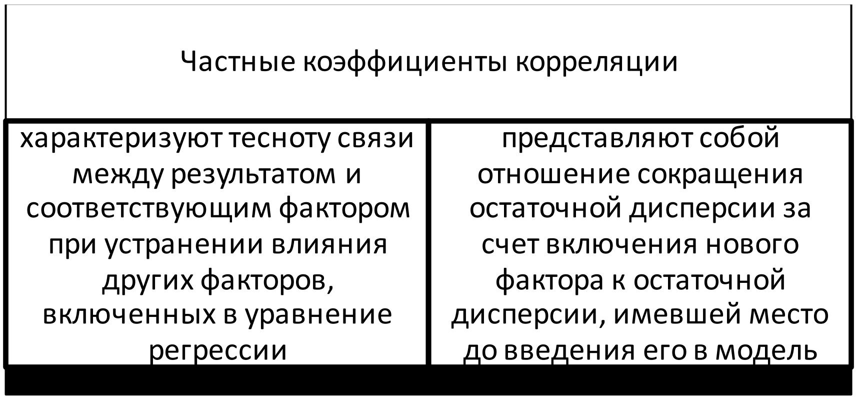

Рис. 4.2. Частные коэффициенты корреляции

$$r_{yx_j \cdot x_1x_2\ldots x_{j-1}x_{j+1}\ldots x_m} = \sqrt{1-\frac{1-R^2{}_{yx_1x_2\ldots x_j\ldots x_m}}{1-R^2{}_{yx_1x_2\ldots x_{j-1}x_{j+1}\ldots x_m}}}$$

где, $R^2yx_1x_2\ldots x_j\ldots x_m$ – множественный коэффициент детерминации всего комплекса факторов с результатом;

$R^2yx_1x_2\ldots x_{j-1,}x_{j+1}\ldots x_m$ – тот же показатель детерминации, но без введения в модель фактора xj.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Коэффициенты парной корреляции называются коэффициентами нулевого порядка.

Коэффициенты частной корреляции первого порядка:

$$r_{yx_1 \cdot x_2} = \beta_{x_1} \cdot \sqrt{\frac{1-r^2{}_{x_1x_2}}{1-r^2{}_{yx_2}}}; r_{yx_2 \cdot x_1} = \beta_{x2} \cdot \sqrt{\frac{1-r^2{}_{x_1x_2}}{1-r^2{}_{yx_1}}}$$

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

$$F = \frac{S^2{}_{ESS}}{S^2{}_{RSS}} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}$$

Мерой для оценки включения дополнительного фактора в модель служит частный F-критерий:

$$F_{x_j} = \frac{R^2y_{x_1\ldots x_j\ldots x_m} - R^2y_{x_1\ldots x_{j-1}x_{j+1}\ldots x_m}}{1-R^2y_{x_1\ldots x_j\ldots x_m}} \cdot \frac{n-m-1}{1}$$

Частный F-критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по модели в целом.

Если наблюдаемое значение частного F-критерия больше критического, то дополнительное включение фактора x_j в модель статистически оправданно и коэффициент b_j статистически значим в предположении, что соответствующий фактор x_j был введен в уравнение множественной регрессии последним.

Оценка значимости коэффициентов регрессии выполняется с помощью t- статистики Стьюдента:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

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

$$\left| t_{b_i} \right| > t_{\alpha/2, n-m-1} \Rightarrow H_1$$

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 $x_1, x_2, \ldots x_m$ в их влияние на зависимую переменную Y.

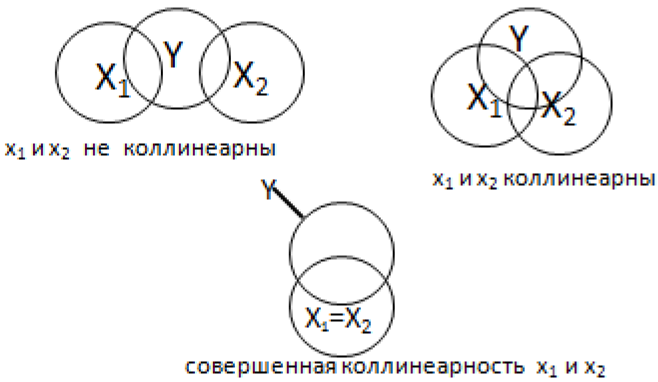

Рис.4.3. Диаграмма Венна

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.

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.

Вопросы для самоконтроля

1. В чем состоит спецификация линейной модели множественной регрессии?
2. Чем скорректированный коэффициент детерминации отличается от обычного?
3. Как определяется статистическая значимость параметров регрессии?
4. Как используется F-статистика в анализе статистической значимости коэффициента детерминации?
5. Сформулируйте требования, предъявляемые к факторам, для включения их в модель множественной регрессии.
6. Назовите методы устранения мультиколлинеарности факторов.
7. Какие коэффициенты используются для оценки сравнительной силы воздействия факторов на результат?
8. Что такое частный F-критерий?
9. Как в линейной модели множественной регрессии, записанной в стандартизованном виде, сравнить факторы по силе их воздействия на результат?
10. Как связаны стандартизованные коэффициенты регрессии с натуральными?

Задания для практики

Задача 1. Уравнение регрессии, построенное по 15 наблюдениям, имеет вид:

$$\begin{array}{lllll} y = & 10 - & 0{,}34x_1 + & 7{,}4x_2 - & ?\,x_3 \\ m_в & (9) & (\;) & (2{,}2) & (4{,}5) \\ t_в & (\;) & (-3{,}4) & (\;) & (-4{,}1) \end{array}$$

Задание: восстановите пропущенные значения и постройте доверительный интервал для b_3 с вероятностью 0,95.

Задача 2. 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 откло-	Парный коэффициент корреляции

		нение	
y	1000	27	ryx1=0,77
x1	420	45	ryx2=0,43
x2	41,5	18	rx1x2=0,38

Задание: постройте уравнение множественной регрессии в стандартизованной и натуральной форме. Определите показатели частной и множественной корреляции. Найдите частные коэффициенты эластичности и сравните их с β-коэффициентами.

Задача 3. Уравнение регрессии в стандартизованных переменных выглядит так: $\hat{t}_y = -0{,}82t_{x_1} + 0{,}65t_{x_2} - 0{,}43t_{x3}$. 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 4. 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: найти регрессию y на x_1 и x_2 в стандартизованной и естественной формах.

Глоссарий

Частное уравнение регрессии характеризует изолированное влияние фактора X_j на результат.

Мультиколлинеарность - это линейная взаимосвязь двух или нескольких объясняющих переменных (x_1, x_2, … x_m).

Частный коэффициент эластичности показывает, на сколько % изменяется в среднем результативный признак Y при изменении фактора Xj на 1%.

Индекс множественной корреляции оценивает тесноту совместного влияния факторов на результативный признак Y.

Бета-коэффициент показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение результативного признака при изменении факторного признака на величину своего среднеквадратического отклонения.

Коэффициент эластичности показывает, на сколько процентов изменяется результативный признак Y при изменении факторного признака X на один процент.

Множественная корреляция – это зависимость между результативным признаком и двумя и более факторными признаками.

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

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11809
http://tulpar.kpfu.ru/mod/resource/view.php?id=11810
http://tulpar.kpfu.ru/mod/resource/view.php?id=11820

2. Бородич С.А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006. – Гл. 6.

3. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010. -Гл. 3.

Лекция 2(2)

Лекция 2(2). Гетероскедастичность и автокорреляция в остатках регрессии

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 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.

Вопросы для изучения:

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

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.

Обнаружение и устранение гетероскедастичности

Методы обнаружения гетероскедастичности: графический анализ остатков; тест Голдфелда-Квандта; тест Спирмена; тест Парка.

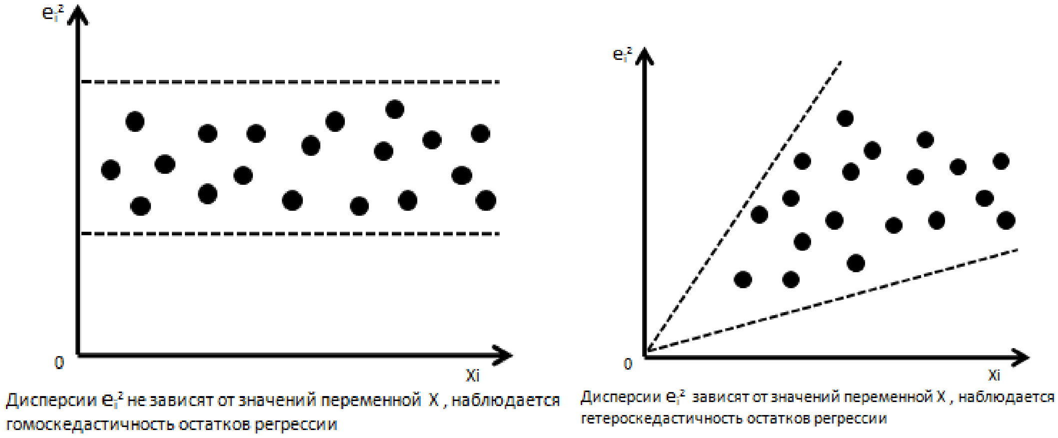

Рис. 5.1. Гомоскедастичность и гетероскедастичность в остатках

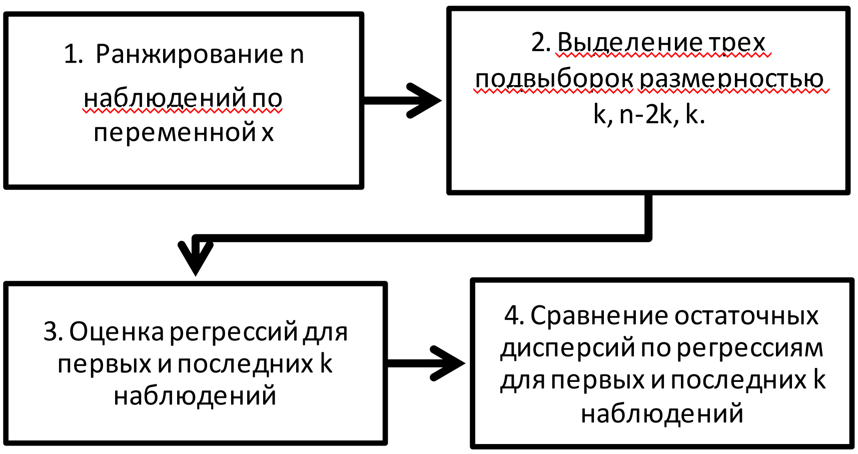

Рис. 5.2. Тест Голдфелда-Квандта

F-статистика для сравнения дисперсий:

$$S^2{}_1=\sum_{i=1}^{k}e_i^{\,2};S^2{}_3=\sum_{i=n-k+1}^{n}e_i^{\,2},$$

$$H_0:S^2{}_3=S^2{}_1(\text{гомоскедастичность})$$

$$H_1:S^2{}_3>S^2{}_1(\text{гетероскедастичность})$$

$$F=\frac{S^2{}_3/(k-m-1)}{S^2{}_1/(k-m-1)},$$

$$F>F_{\alpha,m,k-m-1}\Rightarrow H_1$$

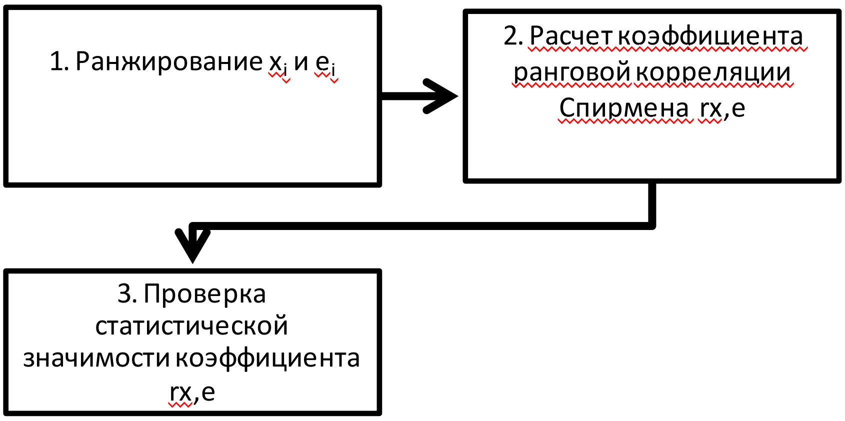

Рис. 5.3. Тест ранговой корреляции Спирмена

t- статистика для проверки значимости $r_{x,e}$:

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

$$H_0:r_{x,e}=0(\text{гомоскедастичность})$$

$$H_1:r_{x,e}\neq 0(\text{гетероскедастичность})$$

$$t=\frac{r_{x,e}\cdot\sqrt{n-2}}{\sqrt{1-r^2{}_{x,e}}},$$

$$t>t_{\alpha,n-2}\Rightarrow H_1$$

Рис.5.4. Тест Парка

t- статистика для проверки значимости коэффициента b:

$$\ln e^2{}_i = a + b\ln x_i + v_i$$

$$H_0 : \beta = 0(\textit{гомоскедастичность})$$

$$H_1 : \beta \neq 0(\textit{гетероскедастичность})$$

$$t = \frac{b}{m_b}$$

$$|t| > t_{\alpha/2, n-2} \Rightarrow H_1$$

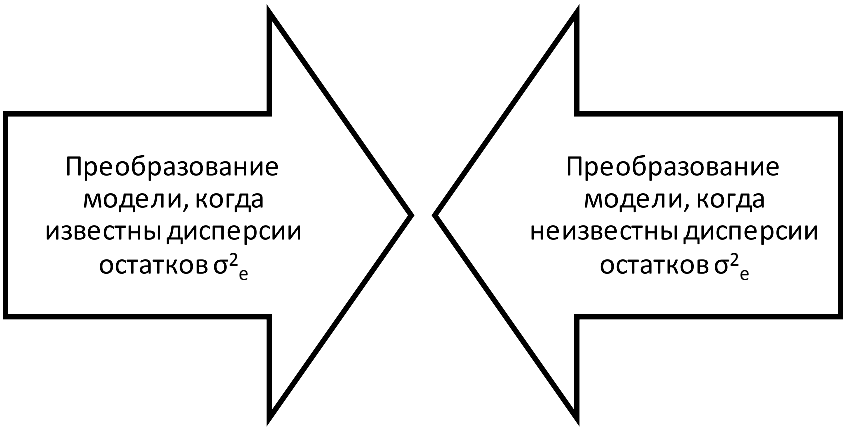

Рис. 5.5. Устранение гетероскедастичности

σ^2_e известны:

$$y = \alpha + \beta \cdot x + \varepsilon$$

$$\frac{y}{\sigma} = \alpha \cdot \frac{1}{\sigma} + \beta \cdot \frac{x}{\sigma} + \frac{\varepsilon}{\sigma}$$

$$y^* = \alpha \cdot z + \beta \cdot x^* + v$$

Метод взвешенных наименьших квадратов:

1. Значения каждой пары наблюдений (xi,yi) делят на известную σi. Тем самым наблюдениям с наименьшими σ^2i придаются наибольшие «веса», а с максимальными дисперсиями – наименьшие «веса». Поэтому наблюдения с меньшими дисперсиями будут более значимыми при оценке коэффициентов регрессии.

2. По МНК для преобразованных значений (1/σi, xi/σi, yi/σi) строится уравнение регрессии с гарантированными качествами оценок.

σ^2_e неизвестны:

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

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

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

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

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

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. 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.

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.

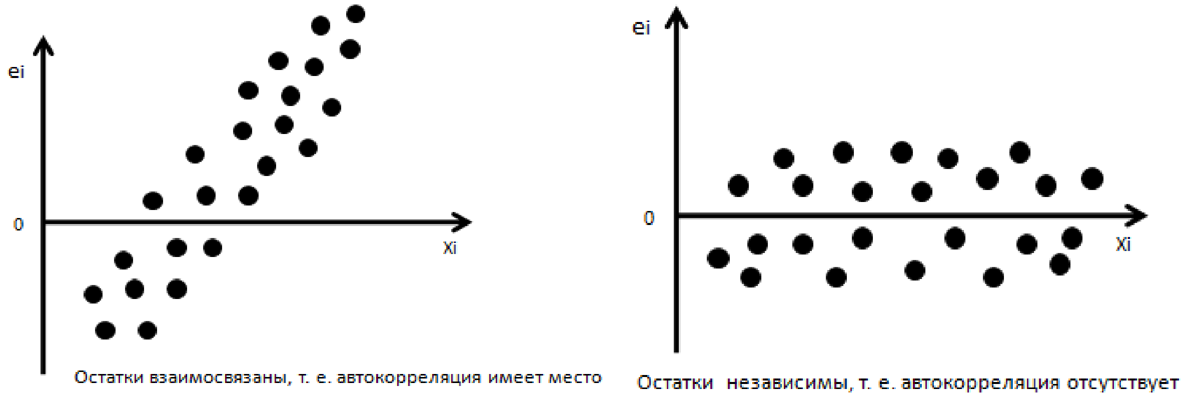

Рис. 5.6. Графический анализ остатков

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. 5.7. Проверка гипотезы об автокорреляции остатков по DW-критерию

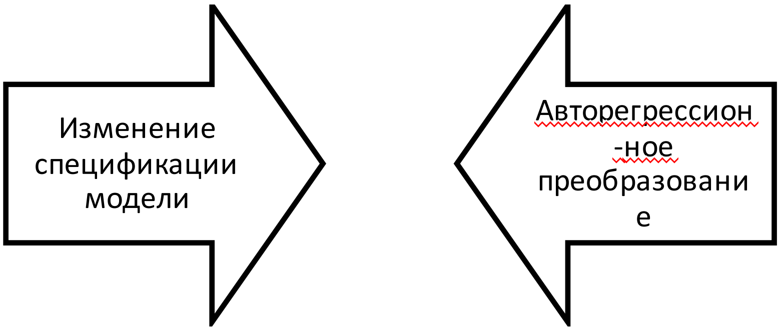

Рис. 5.8. Устранение автокорреляции

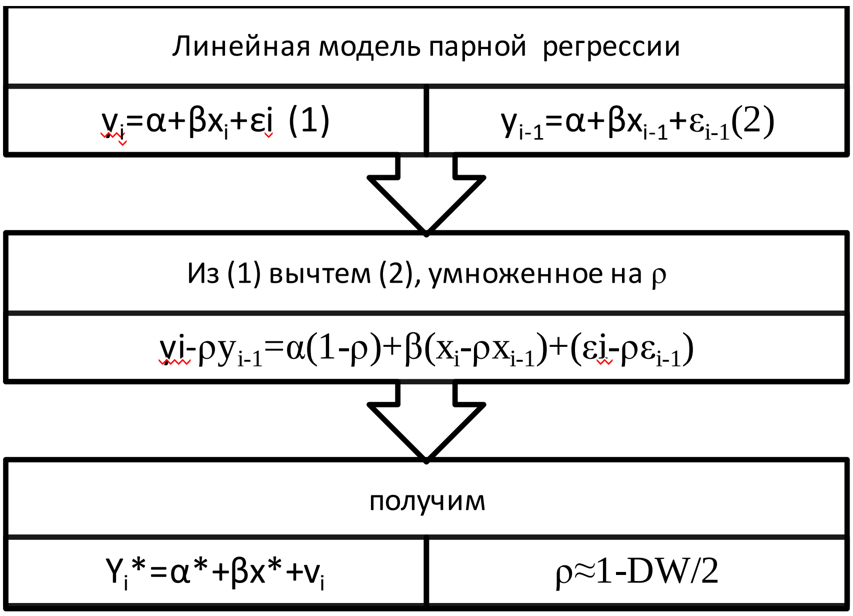

Рис.5.9. Авторегрессионное преобразование

Вопросы для самоконтроля

1. Как проверить наличие гомо- или гетероскедастичности остатков?

2. Как оценивается отсутствие автокорреляции остатков при построении статистической регрессионной модели?

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

4. В чем заключается тест Спирмена?

5. Приведите схему теста Голдфелда-Квандта.

6. Каково предположение теста Парка?

7. В чем суть метода взвешенных наименьшихквадратов?

8. Какие типы преобразований применяются для устранения гетероскедастичности?

9. Что такое автокорреляционная функция?

10. В чем отличие положительной и отрицательной автокорреляции?

11. Какова основная идея метода рядов при обнаружении автокорреляции?

12. Как проводится тест Дарбина-Уотсона?

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

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

15. В чем смысл поправки Прайса-Уинстена?

Задания для практики

Задача 1. Заданы следующие значения остатков линейной модели, соответственные ранжированным значениям фактора x_i :

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 30 странам оценивалась регрессия расходов на образование y от валового национального продукта x по следующим данным:

x	5,67	10,13	11,34	18,88	20,94	22,16	23,83	24,67	27,56	27,57
y	0,34	0,22	0,32	1,23	1,81	1,02	1,27	1,07	0,67	1,25
x	40,15	51,62	57,71	63,03	66,32	66,97	76,88	101,85	115,97	119,49
y	0,75	2,8	4,9	3,5	4,45	1,6	4,26	5,31	6,4	7,15

x	124,15	140,98	153,85	169,38	186,33	211,78	249,72	261,41	395,52	534,97
y	11,22	8,66	5,56	13,11	5,46	4,79	8,92	18,9	15,95	29,9

Задание: построить выборочное уравнение линейной регрессии; проверить наличие гетероскедастичности по критерию Голдфелда-Квандта (уровень значимости $\alpha=0{,}05$); в предположении, что дисперсия отклонений пропорциональна величине валового национального продукта, построить по этим же данным уравнение регрессии по ВМНК; сравнить модели, полученные в п.1 и п.3 и оценить их.

Задача 3. При оценивании модели пространственной выборки с помощью МНК по $n=100$ наблюдениям получено следующее уравнение

$$\tilde{y}=\underset{(0,5)}{12}+\underset{(0,4)}{3{,}43}\,x_1-\underset{(0,1)}{0{,}45}\,x_2,\qquad d=1{,}2.$$

Задание: в скобках указаны стандартные ошибки. С каким из перечисленных выводов следует согласиться: Полученные значения коэффициентов модели с большей вероятностью близки к истинным. Регрессор x_2 может быть незначимым. Так как значение статистики Дарбина-Уотсона d далеко от 2, то следует устранить автокорреляцию остатков.

Задача 4. Имеются данные об урожайности пшеницы y (ц с 1 га) и использовании минеральных удобрений x (кг на 1 га) за 20 лет:

x	36,5	39,1	44,1	45,5	49,0	56,4	66,4	80,9	93,4	109,5
y	18,9	19,9	19,4	19,9	18,5	20,1	21,2	21,9	24,2	24,0
x	123,6	131,6	149,1	157,6	173,6	181,9	193,3	189,0	190,3	188,9
y	23,2	26,5	25,1	29,6	31,7	28,3	31,3	26,9	32,5	29,3

Задание: с помощью теста Дарбина-Уотсона установить наличие или отсутствие автокорреляции на уровне значимости $\alpha=0{,}05$; при наличии автокорреляции определить параметры парной регрессии, используя авторегрессию первого порядка для ошибок регрессии.

Глоссарий

Автокорреляция остатков регрессии – зависимость случайных отклонений ε_i и ε_j друг от друга для $i \neq j$.

Голдфелда Квандта тест – один из наиболее распространенных способов тестирования остатков регрессии на гетероскедастичность

Гомоскедастичность остатков регрессии – постоянство дисперсии случайных отклонений εi.

Дарбина-Уотсона тест – один из наиболее распространенных способов тестирования остатков регрессии на автокорреляцию

Метод наименьших квадратов – один из распространенных способов оценивания параметров регрессии.

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11825

http://tulpar.kpfu.ru/mod/resource/view.php?id=11826

2. Бородич С.А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006. –Гл. 8,9.

3. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 3.

Лекция 3(1)

Лекция 3(1). Фиктивные переменные в регрессионных моделях

Аннотация. Данная тема раскрывает особенности регрессии с фиктивными переменными.

Ключевые слова. Фиктивные переменные, выборка, ANCOVA – модели, тест Чоу.

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.

Вопросы для изучения:

1. Понятие фиктивных переменных.
2. Правило использования фиктивных переменных.
3. ANOVA-модели и ANCOVA-модели.
4. Тест Чоу.

Понятие фиктивных переменных

Исходные статистические данные называют однородными, если все они зарегистрированы при одних и тех же условиях (время года, регион, образование, пол человека). Если же данные объединяют в себе наблюдения, зарегистрированные при различных условиях, то они могут быть неоднородными. В этом случае в модель включается фактор, имеющий два или более качественных уровней.

Влияние качественных факторов иногда приводит к изменению структуры линейных связей в модели (то есть значений коэффициентов *a* и b_i). Построение регрессионной модели по неоднородным данным проводится по одной из двух схем:

- по каждой *регрессионно однородной подвыборке;*

- по объединенной *регрессионно неоднородной выборке* путем введения в модель *фиктивных переменных* (полезно в условиях дефицита исходных данных).

Фиктивные (dummy variables, искусственные, двоичные, структурные) переменные отражают в модели влияние качественного фактора, содержащего атрибутивные признаки двух и более уровней.

Для того, чтобы ввести такие переменные в регрессионную модель, им должны быть присвоены те или иные цифровые метки, то есть качественные переменные необходимо преобразовать в количественные.

Рис. 6.1. Преимущества использования фиктивных переменных

Правило использования фиктивных переменных

Общее правило применения фиктивных переменных: Если качественная переменная имеет k альтернативных значений, то при моделировании используются только (k-1) фиктивных переменных.

Допустим,

$y=a+b*x+\gamma*D_1+ \gamma*D_2+\varepsilon$,

$D_1 = 1$ - мужской пол, $D_1 = 0$ - женский пол;

$D_2 = 1$ - женский пол, $D_2 = 0$ - мужской пол.

Между переменными D_1 и D_2 существует строгая линейная зависимость: $D_2 =1- D_1$, то есть имеет место совершенная мультиколлинеарность, при которой коэффициенты уравнения регрессии однозначно определены быть не могут, так называемая *ловушка фиктивной переменной.*

Значения фиктивной переменной можно изменять на противоположные: D=1- женский пол, D=0 - мужской пол. При этом знак коэффициента γ изменится на противоположный.

Значение качественной переменной, для которого D=0, называется *базовым* или *сравнительным.*

Коэффициент $\boldsymbol{\gamma}$ в модели называется *дифференциальным коэффициентом свободного члена.* Он показывает, на какую величину отличается свободный коэффициент ***a*** при значении D=1, от свободного коэффициента ***a*** при D=0.

ANOVA-модели и ANCOVA-модели

Регрессионные модели, содержащие лишь качественные объясняющие переменные, называются *ANOVA-моделями (моделями дисперсионного анализа).*

$y=a+\gamma*D+\varepsilon$, где

y- заработная плата

D=1- высшее образование ($y=a+\gamma*1= a+\gamma$)

D=0 – нет высшего образования ($y=a+\gamma*0=a$)

Коэффициент *a* определяет среднюю заработную плату при отсутствии высшего образования. Коэффициент $\boldsymbol{\gamma}$ указывает, на какую величину отличаются средние заработные платы при наличии и при отсутствии высшего образования у претендента.

Регрессионные модели, в которых объясняющие переменные носят как количественный, так и качественный характер, называются *ANCOVA-моделями (моделями ковариационного анализа).*

Рис.6.2. Виды Ancova-моделей

Ancova-модель при наличии у фиктивной переменной двух альтернатив:

$y=a+b*x+\gamma*D+\varepsilon$, D=1- лица мужского пола, D=0 – лица женского пола. Ожидаемое потребление кофе при цене *x* будет:

$y=a+b*x+\varepsilon$ для женщины;

$y=a+b*x+\gamma*D+\varepsilon=(a+\gamma)+b*x+\varepsilon$ – для мужчины.

Если γ будет статистически значим по t-статистике, то пол влияет на потребление кофе. При $\gamma>0$ - в пользу мужчин, при $\gamma<0$ – в пользу женщин.

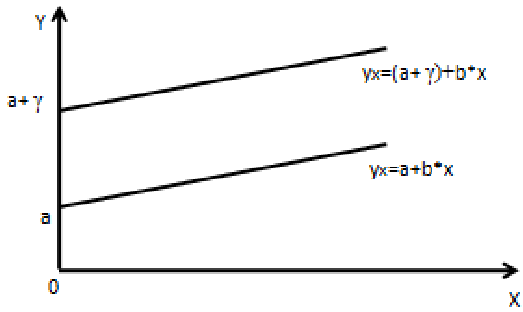

Рис. 6.3. Ancova-модель при наличии у фиктивной переменной двух альтернатив

Ancova-модель при наличии у фиктивной переменной более двух альтернатив:

$y=a+b*x+\gamma_1*D_1+\gamma_2*D_2+\varepsilon$,

Y- расходы на содержание ребенка, X- доходы домохозяйств, $D_1=0$- дошкольник, $D_1=1$- в противоположном случае, $D_2=0$ – дошкольник или младший школьник, $D_2=1$ – в противоположном случае.

Ожидаемые средние расходы при доходах ***x*** будут:

$y=a+b*x$ - на дошкольника;

$y=(a+\gamma_1)+b*x$ –на младшего школьника;

$y=(a+\gamma_1+\gamma_2)+b*x$ – на старшего школьника.

Если γ_1, γ_2 – дифференциальные свободные члены, будут статистически значимы по t-статистике, то возраст ребенка влияет на расходы по его содержанию.

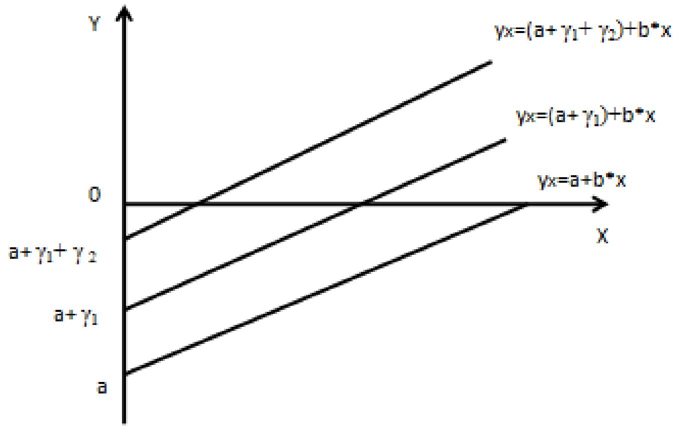

Рис. 6.4. Ancova-модель при наличии у фиктивной переменной более двух альтернатив

Регрессия с одной количественной и двумя качественными переменными:

$y=a+b*x+\gamma_1*D_1+\gamma_2*D_2+\varepsilon$,

y - заработная плата сотрудников фирмы, x- стаж работы, $D_1=0$- женщина, $D_1=1$- мужчина, $D_2=0$ – нет высшего образования, $D_2=1$ – есть высшее образование.

Ожидаемая средняя заработная плата при стаже ***x*** будет:

$y=a+b*x$- для женщины без высшего образования;

$y=(a+\gamma_2)+b*x$ –для женщины с высшим образованием;

$y=(a+\gamma_1)+b*x$ – для мужчины без высшего образования;

$y=(a+\gamma_1+\gamma_2)+b*x$ – для мужчины с высшим образованием.

Если γ_1, γ_2 – дифференциальные свободные члены, будут статистически значимы по t-статистике, то пол сотрудника и его образование влияют на среднюю заработную плату.

Тест Чоу

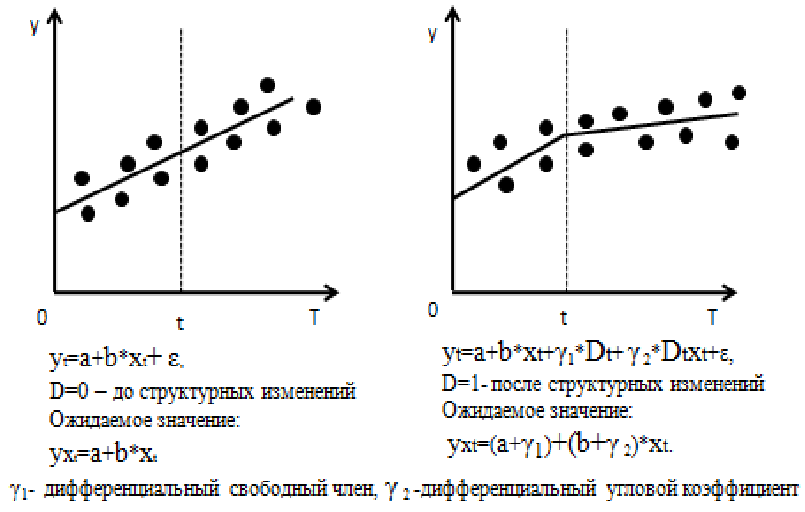

Рис. 6.5. Сравнение двух регрессий

Тест Чоу:

1. Имеется выборка объемом ***n***. Пусть $\boldsymbol{S_0}$ – остаточная сумма квадратов отклонений.

2. Допустим, что выборка объемом ***n*** разбивается на две подвыборки ($\boldsymbol{n_1+n_2=n}$). $\boldsymbol{S_1}$ и $\boldsymbol{S_2}$- суммы квадратов отклонений каждой из подвыборок.

3. Чем сильнее различие в поведении Y для двух подвыборок, тем больше значение $\boldsymbol{S_0}$ будет превосходить $\boldsymbol{S_1+S_2}$. Улучшение качества модели: $\boldsymbol{S_0-(S_1+S_2)}$.

4. $F=\dfrac{S_0-S_1-S_2}{S_1+S_2}\cdot\dfrac{n-2m-2}{m+1}$.

5. Если $F_{набл} < F_{\alpha;m+1;n-2m-2}$, то различие между S_0 и (S_1+S_2) статистически не значимо и нет смысла разбивать уравнение регрессии на части.

Вопросы для самоконтроля

1. Какие статистические данные называют неоднородными?
2. Когда применяются фиктивные переменные?
3. В чем преимущества фиктивных переменных?
4. Как фиктивные переменные включаются в модель регрессии?
5. В чем суть ANOVA-моделей?
6. В чем суть ANCOVA-моделей?
7. В чем состоит правило применения фиктивных переменных?
8. Какой смысл имеет дифференциальный свободный член?
9. Какой смысл имеет дифференциальный угловой коэффициент?
10. В чем особенность моделей с переменной структурой?
11. Какова идея теста Чоу?

Задания для практики

Задача 1. Исследуется зависимость заработной платы y от возраста рабочего x для мужчин и женщин. Оценивание объединенной регрессии ($n=20$) и отдельных регрессий для рабочих-мужчин ($n_1=13$) и рабочих-женщин ($n_2=7$) дали следующие результаты:

Выборка	Оцененное уравнение	R^2	Сумма квадратов остатков
Объединенная	$\tilde{y}=62{,}27+7{,}23x$	0,728	24888
Мужчины	$\tilde{y}=55+7{,}39x$	0,735	18619
Женщины	$\tilde{y}=59{,}43+7{,}3x$	0,712	5658

Задание: улучшилось ли качество регрессии после разделения выборки на части? Найти ответ на уровне значимости $\alpha = 0{,}05$ с использованием критерия Чоу.

Задача 2. При построении линейной зависимости расходов на одежду (y) от располагаемого дохода (x) по выборке для 10 женщин получены следующие суммы квадратов:

$$\sum_{i=1}^{10} x_i = 110,\ \sum_{i=1}^{10} x_i^2 = 1540,\ \sum_{i=1}^{10} y_i = 60,\ \sum_{i=1}^{10} y_i^2 = 448,\ \sum_{i=1}^{10} x_i y_i = 828.$$

Аналогичные вычисления сумм по выборке из 5 мужчин дали:

$$\sum_{i=1}^{5} x_i = 35,\ \sum_{i=1}^{5} x_i^2 = 325,\ \sum_{i=1}^{5} y_i = 15,\ \sum_{i=1}^{5} y_i^2 = 61,\ \sum_{i=1}^{5} x_i y_i = 140.$$

По общей (объединенной) выборке оценена регрессия с использованием фиктивной переменной z (z =1 для мужчин и z =0 для женщин), которая имеет вид:

$$\tilde{y} = \text{-}0{,}06 + 0{,}438x + 0{,}46z .$$

Задание: на уровне $\alpha = 0{,}05$ с использованием теста Чоу проверить гипотезу о том, что функция потребления одна и та же для мужчин и женщин.

Задача 3. На предприятии используются станки трех фирм (А, В, С). Исследуется надежность этих станков. При этом учитывается возраст станка (x, в месяцах) и время безаварийной работы до последней поломки (y, в часах). Выборка из 40 станков дала следующие результаты:

Фирма	А	В	С	А	С	А	В	С	В	А
x	23	30	65	69	75	63	25	75	75	52
y	280	230	112	176	90	176	216	110	45	200
Фирма	В	С	С	В	А	А	С	В	А	А
x	20	70	62	40	66	20	39	25	48	59
y	265	148	150	176	123	245	176	260	236	205

продолжение таблицы

Фирма	А	В	А	С	В	А	С	В	А	В
x	25	69	71	26	45	40	30	69	30	22
y	240	65	115	200	126	225	210	45	260	220
Фирма	В	С	А	В	А	С	В	А	В	А
x	33	48	75	21	56	58	50	37	56	67
y	194	156	100	240	170	116	120	240	88	120

Задание:

1) оценить уравнение регрессии $y = \beta_0 + \beta_1 x + \varepsilon$ без учета различия станков разных фирм;

2) оценить уравнение регрессии, учитывающее различие качества станков разных фирм;

3) сделать вывод о необходимости использования фиктивных переменных в этом случае.

Глоссарий

ANCOVA-модель – это регрессионная модель, в которой объясняющие переменные носят как количественный, так и качественный характер.

Базовое значение качественной переменной имеет цифровую метку ноль: D=0.

Выборка – это часть генеральной совокупности.

Дифференциальный коэффициент свободного члена – это коэффициент перед фиктивной переменной в регрессионной модели. Он показывает, на какую величину отличается свободный коэффициент ***a*** при значении D=1, от свободного коэффициента ***a*** при D=0.

Дифференциальный угловой коэффициент – это коэффициент перед произведением фиктивной переменной и независимой переменной в регрессионной модели. Он показывает, на какую величину отличается коэффициент регрессии ***b*** при значении D=1, от коэффициента регрессии ***b*** при D=0.

Ловушка фиктивной переменной – это состояние совершенной мультиколлинеарности в силу строгой линейной зависимости между переменными D_1 и D_2, при котором коэффициенты уравнения регрессии однозначно определены быть не могут.

Однородные статистические данные – это совокупность данных, зарегистрированных при одних и тех же условиях.

Фиктивные переменные - качественные переменные, преобразованные в количественные с помощью цифровых меток.

Чоу тест – это статистический тест, определяющий целесообразность использования фиктивной переменной.

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11830
http://tulpar.kpfu.ru/mod/resource/view.php?id=11831

2. Бородич С.А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006. –Гл. 11.

3. Эконометрика: учеб. / под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 4.

Лекция 3(2)

Лекция 3(2). Модели с дискретной зависимой переменной

Аннотация. Данная тема раскрывает особенности моделей с дискретной зависимой переменной.

Ключевые слова. Логит-модель, пробит-модель, тест Вальда, тест множителей Лагранжа.

Методические рекомендации по изучению темы

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

•В качестве самостоятельной работы предлагается выполнить практические задания.

Вопросы для изучения:

1. Модели бинарного выбора.
2. Оценивание параметров моделей бинарного выбора.
3. Модели множественного выбора с неупорядоченными альтернативами.
4. Модели множественного выбора с упорядоченными альтернативами.

Модели бинарного выбора

Зависимую переменную, которая принимает несколько значений, называют дискретной. Например, наличие собственного жилья: да – 1, нет – 2. В зависимости от числа альтернатив выделяют модели бинарного и множественного выбора. Чаще применяются модели бинарного выбора. Бинарная переменная принимает лишь два значения: 0 и 1. Например, 1 – занятый, 0 – безработный; 1 – есть мобильный телефон, 0 – нет мобильного телефона. Следовательно, вектор $Y=(y_1, y_2, \ldots, y_n)$ исходных данных будет содержать только дихотомические (бинарные) признаки 0 и 1.

$$P(y_i = 1) = F(x_i'\beta)$$

$$P(y_i = 0) = 1 - F(x_i'\beta)$$

$$0 \le F(\bullet) \le 1$$

Выбор функции $F(\bullet)$ определяет тип бинарной модели. Если используют функцию стандартного нормального распределения,

$$F(u) = \hat{O}(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{z2}{2}} dz$$

то модель бинарного выбора называют пробит-моделью (probit model). Если используют функцию логистического распределения,

$$F(u) = \Lambda(u) = \frac{e^u}{1+e^u}$$

то модель бинарного выбора называют логит-моделью (logit model).

$$P(y_i = 1) = p_i = \frac{1}{1+e^{-(\beta_0+\beta_1 x_i)}} = \frac{1}{1+e^{-z}}$$

$$P(y_i = 0) = 1 - p_i = \frac{1}{1+e^{z_i}}$$

$$\frac{p_i}{1-p_i} = \frac{1+e^{z_i}}{1+e^{-z}} = e^{z_i}$$

$$\ln \frac{p_i}{1-p_i} = z_i = \beta_0 + \beta_1 x_i$$

Оценивание параметров моделей бинарного выбора

Для оценивания параметров β в моделях бинарного выбора обычно используют метод максимального правдоподобия. Общее уравнение правдоподобия:

$$L(\beta) = \prod_{i=1}^{N} p(y_i \ldots)^{1-y_i} \quad P(y_i = 1|x_i;\beta) = F(x_i'\beta),$$

получим, $\log L(\beta) = \sum_{i=1}^{N} y_i \log F(x_i'\beta) + \sum_{i=1}^{N} (1-y_i)\log(1-F(x_i'\beta)).$

Дифференцируя равенство по β, получим уравнение правдоподобия:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{N} \left[\frac{y_i - F(x_i'\beta)}{F(x_i'\beta)(1-F(x_i'\beta))} f(x_i'\beta) \right] x_i = 0, \; f = F'$$

Для логит-модели уравнение упрощается:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{N} \left[y_i - \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \right] x_i = 0$$

Отсюда мы можем найти вероятность того, что y_i=1:

$$\hat{p}_i = \frac{\exp(x_i'\hat{\beta})}{1+\exp(x_i'\hat{\beta})}$$

Уравнение правдоподобия является системой нелинейных (относительно β) уравнений и решается обычно итерационными методами. Для пробит- и логит-моделей данная функция является вогнутой по β, следовательно, решение уравнения правдоподобия дает оценку максимального правдоподобия параметров β_i. Унифицированного, как в линейной регрессии R^2, показателя качества «подгонки» модели не существует. Пусть, log L_f - значение функции правдоподобия исходной модели, log L_c -значение функции правдоподобия той же модели с нулевыми параметрами, но с константой. Чем больше их разность, тем лучше должна быть модель. На этой идее основаны нижеследующие показатели качества модели:

$$EpseudoR^2 = 1 - \frac{1}{1 + 2(\log L_f - \log L_c)/n}$$

$$McFaddenR^2 = 1 - \frac{\log L_f}{\log L_c}$$

Чем больше значение этих показателей, тем лучше модель. Данные показатели редко достигают значений, превышающих 0,5. Для проверки гипотезы о значимости коэффициентов моделей бинарного выбора применяют:

- тест Вальда (Wald test);
- тест множителей Лагранжа (Lagrange multiplier (LM) test);
- отношения правдоподобия (Likelihood ratio (LR) test).

Статистика Вальда имеет распределение χ^2 с числом степеней свободы, равным количеству ограничений в модели. Если наблюдаемое значение превышает критическое для заданного уровня значимости, то нулевая гипотеза о равенстве коэффициентов нулю отклоняется. В качестве аналога F-теста в линейной регрессии о совместной незначимости всех коэффициентов в бинарных моделях используют LR – тест.

$$LR = -2(\log \hat{L}_c - \log \hat{L}_f)$$

LR – тест имеет χ^2 распределение с числом степеней свободы, равным количеству независимых переменных в модели. Если наблюдаемое значение превышает критическое, то нулевая гипотеза о незначимости коэффициентов отклоняется в пользу альтернативной.

Модели множественного выбора с неупорядоченными альтернативами

Модели множественного выбора (multinomial, multi-response models) используются в тех случаях, когда имеется более чем две альтернативы.

Различают: модели с упорядоченными альтернативами (ordered response models); модели с неупорядоченными альтернативами (unordered response models).

Если существует логическое упорядочивание М альтернатив, то может использоваться дискретная модель с упорядоченными альтернативами. Эта модель основывается на предположении о существовании одной ненаблюдаемой латентной переменной Y_i^* : $y^*{}_i = x_i'\beta + \varepsilon_i$

Стандартное нормальное распределение остатков дает упорядоченную probit-модель (ordered probit model). Логистическое распределение остатков дает упорядоченную logit-модель (ordered logit model).

Для случая трех альтернатив:

$$P(y_i = 1|x_i) = P(y_i^* \le 0|x_i) = \hat{O}(-x_i'\beta)$$

$$P(y_i = 3|x_i) = P(y_i^* > \gamma|x_i) = 1 - \hat{O}(\gamma - x_i'\beta)$$

$$P(y_i = 2|x_i) = \hat{O}(\gamma - x_i'\beta) - \hat{O}(-x_i'\beta)$$

Для случая М вариантов выбора:

$$P(y_i = 0|x_i) = P(y_i^* \le 0|x_i) = \hat{O}(-x_i'\beta)$$

$$P(y_i = 1|x_i) = \hat{O}(\gamma_1 - x_i'\beta) - \hat{O}(-x_i'\beta)$$

$$P(y_i = 2|x_i) = \hat{O}(\gamma_2 - x_i'\beta) - \hat{O}(\gamma_1 - x_i'\beta)$$

$$\ldots$$

$$P(y_i = M|x_i) = P(y_i^* > \gamma|x_i) = 1 - \hat{O}(\gamma_{M-1} - x_i'\beta)$$

Оценивание осуществляется при помощи метода максимального правдоподобия, где перечисленные вероятности включены в функцию правдоподобия.

$$\log L(\beta,\gamma) = \sum_{i:y_i=0} \log(\Pr(y_i = 0|x_i,\beta,\gamma)) + \sum_{i:y_i=1} \log(\Pr(y_i = 1|x_i,\beta,\gamma)) +$$
$$+ \ldots + \sum_{i:y_i=M} \log(\Pr(y_i = M|x_i,\beta,\gamma)).$$

Модели множественного выбора с упорядоченными альтернативами

В некоторых случаях не существует естественного упорядочивания между альтернативами. Например, при моделировании способа передвижения (автобус, поезд, машина, велосипед, пешком). Предполагается существование случайной полезности, которая влияет на выбор альтернатив.

Случайные полезности являются линейными функциями от наблюдаемых характеристик и имеют аддитивно-разделяемую структуру.

Полезность:

$$U_{ij} = \mu_{ij} + \varepsilon_{ij}$$

μ_{ij} - неслучайная функция наблюдаемых неизвестных параметров;

ε_{ij} – ненаблюдаемый остаточный член.

$$P(y_i = j) = P(U_{ij} = \max(U_{i1},...U_{iM})) =$$
$$= P(\mu_{ij} + \varepsilon_{ij} > \max_{k=1,...,J,k\neq j}(\mu_{ik} + \varepsilon_{ik}).$$

Предположим, что все ε_{ij} взаимно независимы и распределены по закону распределения Вейбулла.

$$P(y_i = j) = \frac{\exp(\mu_{ij})}{\exp(\mu_{i1}) + \exp(\mu_{i2}) + ... + \exp(\mu_{iM})}$$
$$0 \le P(y_i = j) \le 1$$
$$\sum_{j=1}^{M} P(y_i = j) = 1$$

Один из уровней полезности принимают равным нулю (μ_{i1}=0) и полагают, что μ_{ij} является линейной функцией от наблюдаемых переменных:

$$\mu_{ij} = x'_{ij}\beta$$
$$P(y_i = j) = \frac{\exp(x'_{ij}b)}{1 + \exp(x'_{i2}b) + ... + \exp(x'_{iM}b)},\ j = 1,2,...M.$$

Данное выражение представляет собой logit – модель с множественными альтернативами (Multinomial Logit Model (MNL) или Independet Logit Model).

Вопросы для самоконтроля

1. В каких ситуациях фиктивная переменная используется в качестве зависимой переменной?

2. Какие законы распределения чаще всего используются в моделях бинарного выбора?

3. В чем суть логит-модели?

4. В чем суть пробит-модели?

5. Какова интерпретация коэффициентов моделей бинарного выбора?

6. Как осуществляется проверка значимости коэффициентов в модели бинарного выбора?

7. Как получить прогноз вероятности по логит-модели?

8. Как получить прогноз вероятности по пробит-модели?

9. Можно ли рассчитать по логит-модели коэффициент детерминации?

10. Какие существуют варианты постановки моделей множественного выбора?

11. В чем отличие моделей упорядоченного и неупорядоченного выбора?

Задания для практики

Задача 1. Банк исследует вероятность невозвращения потребительского кредита (y=1 – заемщик кредит возвращает, y=0 – не возвращает), используя два фактора: x_1 – сумма займа, x_2 - среднемесячный доход заемщика. По логит-модели:

$$P_i = \frac{1}{1+e^{-(5-0,6(x_1/x_2))}},$$

оцените вероятность невозвращения кредита при покупке на сумму 40 тыс. руб. и доходе 10 тыс. руб. Повторите расчет при стоимости покупки в 50 тыс. руб. и доходе 5 тыс. руб. Дайте рекомендацию банку о пороговом соотношении суммы займа и среднемесячного дохода, чтобы предсказанная по модели доля просроченных кредитов не превышала 5%.

Задача 2. Исследуется вопрос о наличии собственного дома ($Y=1$, если дом имеется; $Y=0$, если дома нет) в зависимости от совокупного дохода семьи (X). Выборка из 40 семей дала следующие результаты:

Семья	1	2	3	4	5	6	7	8	9	10
X	10	20	22	18	9	15	25	30	40	16
Y	0	1	1	0	0	0	1	1	1	0

Семья	11	12	13	14	15	16	17	18	19	20
X	12	8	20	19	30	50	37	28	45	38
Y	0	0	1	0	1	1	1	1	1	1

Семья	21	22	23	24	25	26	27	28	29	30
X	30	12	16	27	19	15	32	18	43	13
Y	1	0	0	1	0	0	1	0	1	0

Семья	31	32	33	34	35	36	37	38	39	40
X	22	14	10	17	36	45	14	22	41	34

Y	1	0	0	0	1	1	0	1	1	1

Задание:

1) построить линейную вероятностную модель;
2) оценить качество построенной модели;
3) оценить вероятность того, что при доходе, равном 18, семья имеет дом.

Задача 3. При найме на работу претендентам предлагается выполнить тестовое задание, X- стаж работы, мес., Y – результаты теста.

X	7	15	16	15	8	4	18	2	22	6	30	1
Y	0	0	0	1	1	0	0	0	1	0	1	0
X	30	5	20	13	9	32	4	13	9	4	28	22
Y	1	0	1	0	0	1	0	1	0	0	1	1

Задание: проверить, зависит ли успешное выполнение теста от стажа работы, построить логит- и пробит-модели, оценить значимость уравнений на уровне значимости α=0,05 и ответить на вопросы:

1) чему равна вероятность успешного выполнения задания при стаже в 1 месяц, 5 месяцев, 15 месяцев?

2) на какую величину повышает вероятность выполнения задания каждый следующий месяц при стаже 1, 5, 15 месяцев?

Задача 4. В следующей таблице представлены данные о количестве семей (N), имеющих определенный уровень дохода (X), и количестве семей (n), имеющих частные дома:

X	10	15	20	25	30	35	40	45	50	55	60
N	35	45	60	80	100	130	90	65	50	30	15
n	5	10	18	30	45	60	55	45	38	24	13

Задание: Оценить logit-модель по МНК.

Глоссарий

Дискретная зависимая переменная – это переменная, которая принимает несколько альтернативных значений.

Логит-модель основана на использовании функции логистического распределения.

Метод максимального правдоподобия – один из способов оценивания параметров регрессии, в частности, в моделях с дискретной зависимой переменной.

Модели бинарного выбора содержат зависимую переменную, которая принимает лишь два альтернативных значения, обозначаемых цифровыми метками: 0 и 1.

Модели множественного выбора содержат зависимую переменную, которая принимает несколько упорядоченных или неупорядоченных альтернативных значений.

Пробит-модель основана на использовании функции стандартного нормального распределения.

Использованные информационные ресурсы

1. Бородич С.А. Эконометрика: учебное пособие. - Мн.: Новое знание, 2006. –Гл. 11.

2. Эконометрика: учеб. / под ред. В. С. Мхитаряна.- М.: Проспект, 2008. - Гл. 8.

3. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 5.

Лекция 3(3)

Лекция 3(3). Нелинейные модели регрессии и их линеаризация

Аннотация. Данная тема раскрывает виды и способы оценивания нелинейных моделей регрессии.

Ключевые слова. Нелинейная регрессия, индекс корреляции, подход Бокса-Кокса.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где даются общие представления по теме.
- В качестве самостоятельной работы предлагается выполнить практические задания.

Вопросы для изучения:

1. Классы и виды нелинейных регрессий. Индекс корреляции.
2. Линеаризация нелинейных моделей.
3. Выбор формы модели. Подбор линеаризующего преобразования (подход Бокса-Кокса).

Классы и виды нелинейных регрессий. Индекс корреляции

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью нелинейных функций.

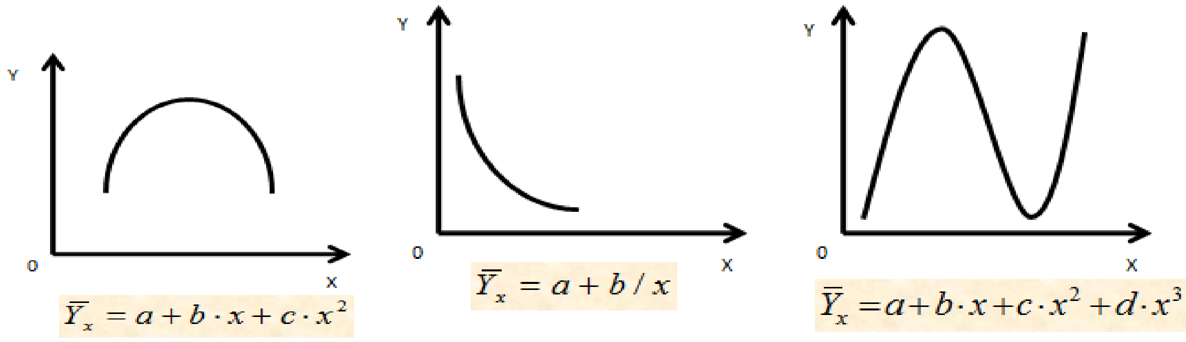

Рис. 7. 1. Регрессии, нелинейные относительно переменных

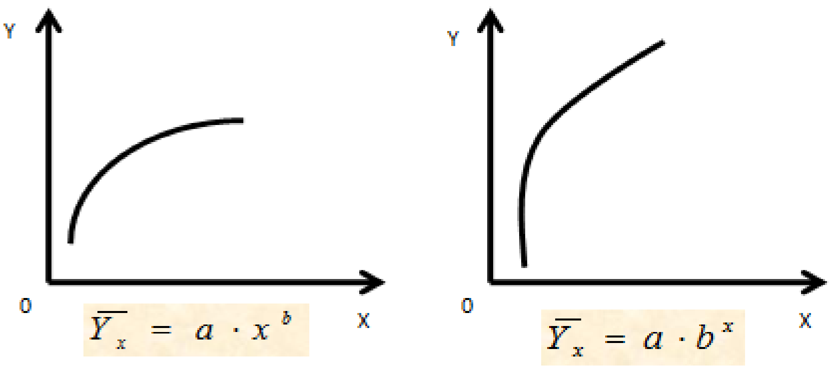

Рис.7.2. Регрессии, нелинейные по оцениваемым параметрам

Нелинейная модель, внутренне линейная, с помощью преобразований может быть приведена к линейному виду. Нелинейная модель, внутренне нелинейная, не может быть сведена к линейной функции.

Для измерения тесноты связи между переменными в нелинейных регрессиях применяется индекс корреляции:

$$R = \sqrt{1 - \frac{\sum (y - y_x)^2}{\sum (y - \bar{y})^2}}$$

Линеаризация нелинейных моделей

Способы линеаризации		
Замена переменных	Логарифмирование обеих частей уравнения	Комбинированный

Рис. 7.3. Способы линеаризации

Замена переменных заключается в замене нелинейных объясняющих переменных новыми линейными переменными и сведении нелинейной регрессии к линейной. Например, полиномиальная модель:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_k x^k + \varepsilon$$
$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + ... + \beta_k z_k + \varepsilon$$

Например, гиперболическая модель:

$$y = \beta_0 + \frac{\beta_1}{x} + \varepsilon$$
$$y = \beta_0 + \beta_1 z + \varepsilon$$
$$z_1 = \frac{1}{x_1}; z_2 = \frac{1}{x_2}; ... z_n = \frac{1}{x_n}$$

Например, кривая Филлипса (равносторонняя гипербола), где

x- норма безработицы, y – процент прироста заработной платы:

$$y = \frac{1}{\beta_0 + \beta_1 x + \varepsilon}, \left(-\frac{\beta_0}{\beta_1} < x < \infty \right)$$
$$z = \beta_0 + \beta_1 x + \varepsilon$$
$$z_1 = \frac{1}{y_1}, z_2 = \frac{1}{y_2}, ... z_3 = \frac{1}{y_n}$$

Например, кривая Энгеля, где x- доход потребителей, y – спрос на определенный вид товаров или услуг

$$y = \frac{x}{\beta_0 x + \beta_1 + \varepsilon x}, \left(-\frac{\beta_0}{\beta_1} < x < \infty \right)$$
$$z = \beta_0 + \beta_1 z' + \varepsilon$$
$$z_1 = \frac{1}{y_1}, z_2 = \frac{1}{y_2}, ... z_3 = \frac{1}{y_n}$$
$$z'_1 = \frac{1}{x_1}, z'_2 = \frac{1}{x_2}, ... z'_3 = \frac{1}{x_n}$$

Например, полулогарифмические модели:

1) $\ln y = \beta_0 + \beta_1 x + \varepsilon$

Такие модели обычно используются в тех случаях, когда необходимо исследовать зависимость темпа роста или прироста экономических показателей:

- прирост объема выпуска от процентного увеличения затрат ресурсов;
- прирост бюджетного дефицита от темпа роста ВНП;
- темп роста инфляции от объема денежной массы.

2) $y = \beta_0 + \beta_1 \ln x + \varepsilon \qquad z = \ln y; z = \beta_0 + \beta_1 x + \varepsilon$

Используется обычно в тех случаях, когда необходимо исследовать, как процентное изменение независимой переменной влияет на абсолютное изменение зависимой переменной: влияние относительного (процентного) увеличения денежной массы на абсолютное изменение ВНП.

$$z = \ln x; y = \beta_0 + \beta_1 z + \varepsilon$$

<u>Логарифмирование обеих частей уравнения</u> применяется обычно, когда мультипликативную модель необходимо привести к линейному виду.

Например, степенные модели:

$$y = \beta_0 x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \ldots \cdot x_p^{\beta_k} \cdot \varepsilon$$

$$y' = \ln y; x_j' = \ln x_j; \varepsilon' = \ln \varepsilon (j = 1,2,\ldots k)$$

$$y' = \beta_0' + \beta_1 x_1' + \ldots + \beta_k x_k' + \varepsilon, \text{ãäå} \beta_0' = \ln \beta_0$$

К классу степенных функций относятся: кривые спроса и предложения, производственная функция Кобба-Дугласа, кривые освоения для характеристики связи между трудоемкостью продукции и масштабами производства в период освоения и выпуска нового вида изделий, зависимость валового национального дохода от уровня занятости.

Например, показательные (экспоненциальные) модели. Широкий класс экономических показателей характеризуется приблизительно постоянным темпом относительного прироста во времени. Этому соответствует следующая форма зависимости показателя Y от времени X:

$$y = \beta_0 e^{\beta_1 x + \varepsilon}, (e = 2{,}7182818).$$

$$y' = \ln y$$

$$y' = \beta_0' + \beta_1 x + \varepsilon, \beta_0' = \ln \beta_0$$

$$b_0 = e^{\beta_0'}$$

$$y = \beta_0 e^{\beta_1 \frac{1}{x} + \varepsilon}, (e = 2{,}7182818).$$

$$y' = \ln y, x' = \frac{1}{x}$$

$$y' = \beta_0' + \beta_1 x' + \varepsilon, \beta_0' = \ln \beta_0$$

$$b_0 = e^{\beta_0'}$$

Например, логистическая кривая. Применяется для описания поведения показателей, имеющих определенные «уровни насыщения»: зависимость спроса на товар Y от дохода X, развитие производства новых товаров, рост численности населения (впервые применил А. Кетле (1796-1874)).

$$y = \frac{1}{\beta_0 + \beta_1 e^{-x} + \varepsilon}, (-\infty < x < \infty).$$

$$y' = \frac{1}{y}, x' = e^{-x}$$

$$y' = \beta_0 + \beta_1 x' + \varepsilon$$

Например, логлинейная модель. Используется в банковском и финансовом анализе. Y_0 – начальная величина переменной Y (первоначальная сумма вклада), r – сложный темп прироста величины Y (процентная ставка); Y_t – значение величины Y в момент времени t (вклад в банке в момент времени t).

$$\ln y_t = \ln y_0 + t\ln(1+r) + \ln\varepsilon$$
$$\ln y_0 = \beta_0, \ln(1+r) = \beta_1, \ln\varepsilon_t = \varepsilon_t'$$
$$\ln y_t = \beta_0 + \beta_1 t + \varepsilon_t'$$

Величина коэффициента эластичности показывает, на сколько процентов изменится результативный признак Y, если факторный признак изменится на 1 %:

$$\acute{Y} = f'(x)\frac{x}{y}$$

Расчет коэффициента эластичности:

$$y = a + bx + \varepsilon \to \frac{bx}{a+bx}$$

$$y = a + bx + cx^2 + \varepsilon \to \frac{(b+2cx)\cdot x}{a+bx+cx^2}$$

$$y = a + \frac{b}{x} + \varepsilon \to \frac{-b}{ax+b}$$

$$y = ab^x\varepsilon \to x\cdot\ln b$$

$$y = ax^b\varepsilon \to b$$

$$y = a + b\ln x + \varepsilon \to \frac{b}{a+b\ln x}$$

$$y = \frac{a}{1+b\cdot e^{-cx+\varepsilon}} \to \frac{c\cdot x}{\frac{1}{b}\cdot e^{cx}+1}$$

$$y = \frac{1}{a+bx+\varepsilon} \to \frac{b\cdot x}{a+b\cdot x}$$

Выбор формы модели. Подбор линеаризующего преобразования (подход Бокса-Кокса)

Выбор модели не всегда осуществляется однозначно, и в дальнейшем требуется сравнивать модель как с теоретическими, так и с эмпирическими данными, совершенствовать ее. При определении качества модели обычно анализируются следующие параметры: скорректированный коэффициент детерминации; t – статистики; статистика Дарбина-Уотсона (DW); согласованность знаков коэффициентов с экономической теорией; прогнозные качества (ошибки) модели. Английские статистики Г. Бокс и Д. Кокс предложили формализованную процедуру подбора линеаризующего преобразования. Их метод основан на предположении, что искомое преобразование принадлежит определенному однопараметрическому семейству преобразований вида

$\widetilde{y}_i(\lambda)=\dfrac{y_i^{\lambda}-1}{\lambda}, \widetilde{x}_i(\lambda)=\dfrac{(x_{ij})^{\lambda}-1}{\lambda}, i=1,2,...n.(1)$

Гипотеза Г. Бокса и Д. Кокса: Существует такое вещественное (положительное или отрицательное) число, что один из двух нижеследующих вариантов представления искомой регрессионной зависимости между наблюдаемыми переменными будет удовлетворять всем требованиям нормальной классической линейной модели множественной регрессии.

$\widetilde{y}_i(\lambda)=\beta_0+\beta_1 x_{i1}+...+\beta_k x_{ik}+\varepsilon, i=1,2,...,n(2)$

$\widetilde{y}_i(\lambda)=\beta_0+\beta_1 x_{i1}(\lambda^*)+...+\beta_k x_{ik}(\lambda^*)+\varepsilon, i=1,2,...,n(3)$

Замечание 1. Преобразования вида (1) применяются обычно к переменным, принимающим только положительные значения.
Замечание 2. При $\lambda=1$ модели (2) и (3) являются линейными относительно Y и X.
При $\lambda=0$ между Y и X наблюдается степенная зависимость.

Оценка неизвестного значения параметра λ сводится к оценке параметра в формулах (1) по имеющимся в распоряжении исходным статистическим данным методом максимального правдоподобия.

Вопросы для самоконтроля

1. Назовите классы и виды нелинейных регрессий.

2. Перечислите все виды моделей, нелинейных относительно: а)включаемых переменных; б) оцениваемых параметров.

3. Какие преобразования используются для линеаризации нелинейных моделей?

4. Какой нелинейной функцией может быть заменена парабола второй степени, если не наблюдается смена направленности связи признаков?

5. Чем отличается применение МНК к моделям, нелинейным относительно включаемых переменных, от применения к моделям, нелинейным по оцениваемым параметрам?

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

7. Назовите показатели корреляции, используемые при нелинейных соотношениях рассматриваемых признаков.

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

9. Приведите примеры использования логарифмических регрессионных моделей. Каков смысл коэффициентов регрессии в таких моделях?

10. Приведите примеры использования обратных и степенных моделей.

11. Как интерпретируются коэффициенты регрессии в модели потребления?

12. Какой смысл приобретает сумма коэффициентов регрессии в производственных функциях и что означает, когда такая сумма коэффициентов регрессии больше единицы?

Задания для практики

Задача 1. Изучалась зависимость вида $y = b_0 x^{b_1}$. Для преобразованных в логарифмах переменных получены следующие данные:

$$\sum x_i' = 8{,}2370;\ \sum y_i' = 3{,}9310;\ \sum x_i' y_i' = 4{,}2087;\ \sum x_i'^2 = 9{,}2334;\ n = 10.$$

Задание: найти параметры b_0, b_1.

Задача 2. Анализируется прибыль предприятия Y (млн. долл.) в зависимости от расходов на рекламу X (млн. долл.). По наблюдениям за 9 лет получены следующие данные:

Y	5	7	13	15	20	25	22	20	17
X	0,8	1,0	1,8	2,5	4,0	5,7	7,5	8,3	8,8

Задание:

1) постройте корреляционное поле и выдвиньте предположение о формуле зависимости между рассматриваемыми показателями;

2) оцените по МНК коэффициенты линейной регрессии;

3) оцените качество построенной модели;

4) оцените по МНК коэффициенты квадратичной регрессии;

5) оцените качество построенной модели. Какую из моделей следует предпочесть?

Задача 3. Для трех видов продукции A, B, C модель зависимости удельных постоянных расходов от объема выпускаемой продукции выглядит следующим образом:

$$y_A = 600;$$

$$y_B = 80 + 0{,}7x;$$

$$y_C = 40x^{0{,}5}.$$

Задание:

1) определить коэффициенты эластичности по каждому виду продукции;

2) сравнить при $x = 1000$ эластичность затрат для продукции B и C;

3) определить, каким должен быть объем выпускаемой продукции, чтобы коэффициенты эластичности для продукции B и C были равны.

Задача 4. При исследовании спроса на телевизоры марки Т, аналитический отдел компании АВС по данным, собранным по 19 торговым точкам компании, выявил следующую зависимость:

$$\ln y = 10{,}5 - 0{,}8 \ln x + \varepsilon$$

(2,5) (-4,0)

где: Y - объем продаж телевизоров в отдельной торговой точке, X - средняя цена телевизора в данной торговой точке. В скобках приведены фактические значения $t-$критерия.

Задание: до проведения этого исследования администрация компании предполагала, что эластичность спроса по цене для телевизоров марки Т составляет $\bar{Э}$ = - 0,9. Подтвердилось ли предположение администрации результатами исследования?

Задача 5. По группе из 10 заводов, производящих однородную продукцию, получено уравнение регрессии себестоимости единицы продукции y (тыс. руб.) от уровня технической оснащенности x (тыс. руб.)

$$\tilde{y} = 20 + \frac{700}{x}.$$

Доля остаточной дисперсии в общей составила 0,19.

Задание:

1) определить коэффициент эластичности, предполагая, что стоимость активных производственных фондов составляет 200 тыс. руб.;

2) вычислить индекс корреляции;

3) оценить значимость уравнения регрессии с помощью $F-$критерия.

Глоссарий

Внутренне линейная нелинейная модель с помощью преобразований может быть приведена к линейному виду.

Внутренне нелинейная модель не может быть сведена к линейной функции.

Индекс корреляции – показатель тесноты статистической взаимосвязи, выраженной в нелинейной форме.

Коэффициент эластичности – показатель, который измеряет, на сколько процентов изменится результативный признак Y, если факторный признак изменится на 1 %.

Линеаризация – приведение нелинейных моделей регрессии к линейному виду путем замены, логарифмирования переменных или комбинированным способом с целью применения МНК.

Бокса-Кокса подход - формализованная процедура подбора линеаризующего преобразования.

Использованные информационные ресурсы

1. Бородич С.А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006. –Гл. 7.

2. Эконометрика: учеб. / под ред. В. С. Мхитаряна.- М.: Проспект, 2008. - Гл. 7.

Лекция 4(1)

Лекция 4(1). Модели одномерных временных рядов

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.

Вопросы для изучения:

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

Понятие временного ряда и его основные компоненты

Временной ряд - это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени (y_t).

Модели, построенные по временным рядам, называются моделями временных рядов. Параметры таких моделей оцениваются специальными методами, разработанными на основе традиционных методов регрессионного анализа.

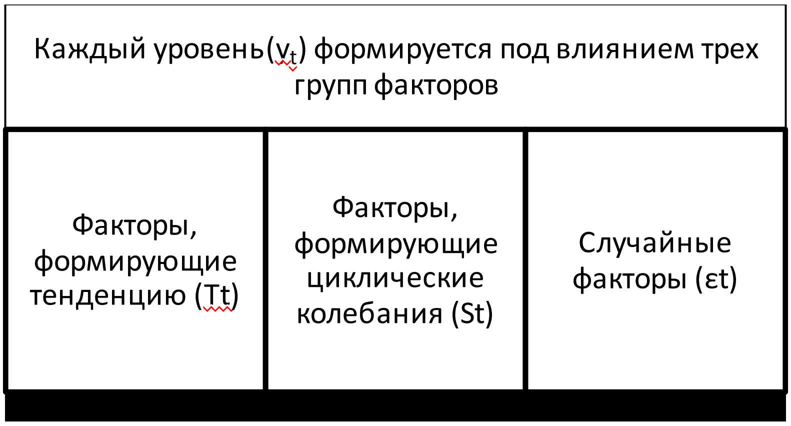

Рис. 8.1. Три компоненты временного ряда

Реальные данные чаще всего содержат все три компоненты.

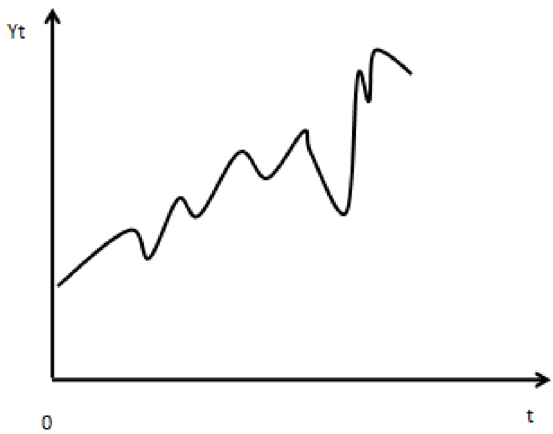

Рис. 8.2. Одновременное присутствие компонент временного ряда

Модель, в которой временной ряд представлен как сумма его компонент, называется аддитивной моделью временного ряда: $y_t = T_t + S_t + e_t$.

Модель, в которой временной ряд представлен как произведение его компонент, называется мультипликативной моделью временного ряда: $y_t = T_t \cdot S_t \cdot e_t$.

Основная задача эконометрического исследования временного ряда – выявление и количественное измерение тенденции, циклической и случайной компонент, с тем, чтобы использовать информацию для получения прогнозных оценок или при построении моделей взаимосвязи двух или более временных рядов.

При наличии тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих значений. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Число периодов, по которым рассчитывается коэффициент автокорреляции, называется лагом. Максимальный лаг должен быть не больше n/4. Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют автокорреляционной функцией временного ряда. График зависимости ее значений от величины лага называется коррелограммой.

Коэффициент автокорреляции уровней ряда первого порядка:

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

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

Коэффициент автокорреляции уровней ряда второго порядка:

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

$$\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}; \bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}$$

Свойства коэффициента автокорреляции: характеризует тесноту только линейной связи текущего и предыдущего уровней ряда; по знаку коэффициента нельзя делать вывод о возрастающей или убывающей тенденции в уровнях.

При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

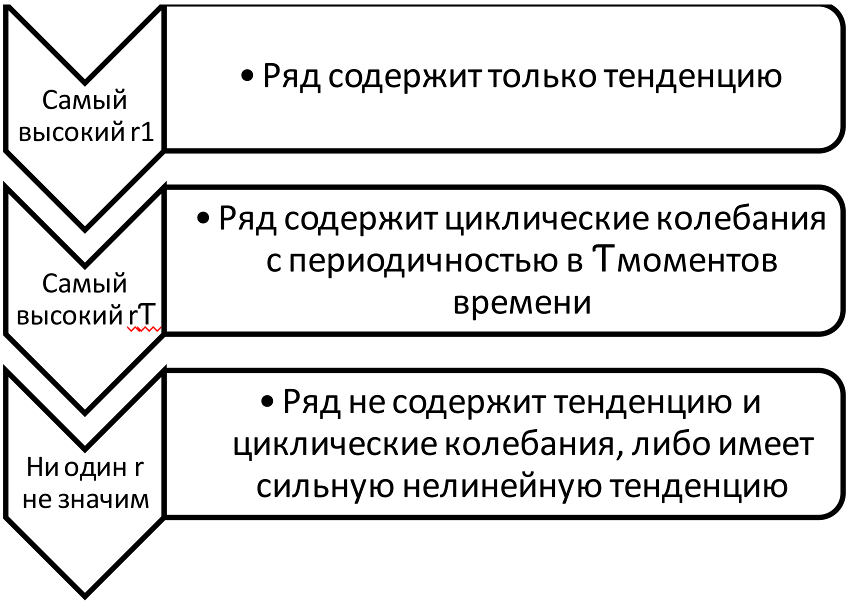

Рис. 8.3. Выводы о структуре временного ряда

Методы выявления основной тенденции временного ряда: сглаживание или механическое выравнивание уровней ряда; аналитическое выравнивание уровней ряда.

Типы трендов:

- линейный тренд: $\hat{y}_t = a + b \cdot t$;
- гипербола: $\hat{y}_t = a + b / t$;
- экспонента: $\hat{y}_t = e^{a+b\cdot t^{l}}$;
- степенной тренд: $\hat{y}_t = a \cdot t^b$;
- парабола k-го порядка: $\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + \ldots + b_k \cdot t^k$.

Приемы выявления типа тенденции: графически; по абсолютным приростам и темпам роста сглаженных уровней; метод последовательных разностей; сравнительная оценка остаточной суммы квадратов и характеристик качества регрессии.

Построение аддитивной модели

Рис. 8.4. Анализ структуры временного ряда

Как выбрать тип модели?

Рис. 8.5. Выбор модели временного ряда

Рис. 8.6. Особенность сезонной компоненты

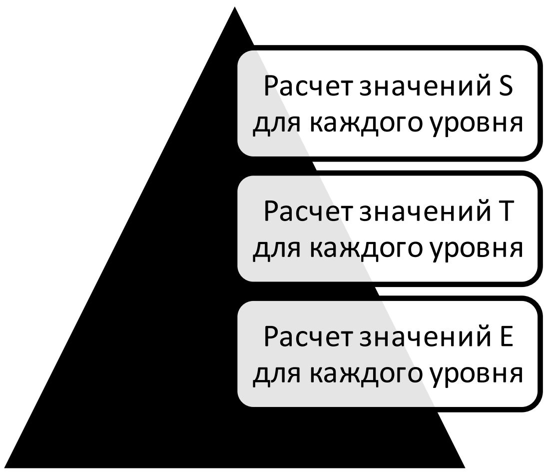

Рис. 8.7. Процесс построения модели

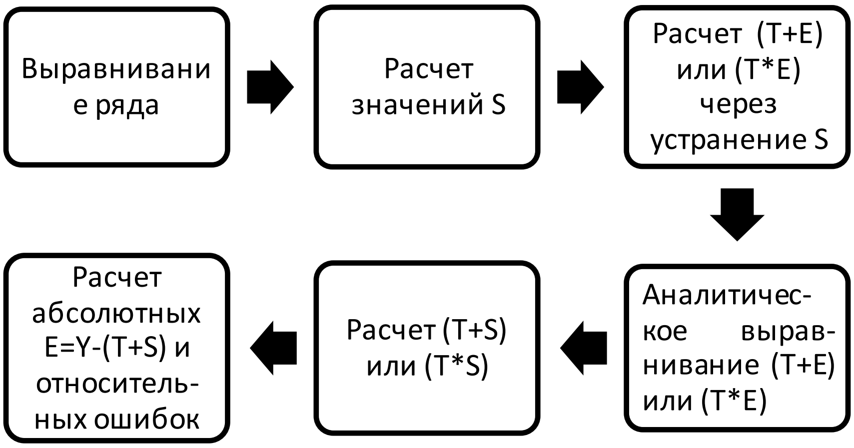

Рис. 8.8. Этапы построения модели

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 Т 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 Т для каждого момента времени по уравнению тренда.

5 шаг. *Расчет значений T+S.* Умножим уровни Т 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 для самоконтроля

1. Из каких основных компонентов складывается уровень временного ряда y_t?

2. Как можно выявить автокорреляцию в остатках ?

3. В каких пределах изменяется статистика Дарбина – Уотсона?

4. Как записывают аддитивную модель временного ряда ?

5. Когда применяют мультипликативную модель временного ряда?

5. В чем суть коэффициента автокорреляции временного ряда ?

6. Что называют лагом во временных рядах ?

7. В чем заключается метод аналитического выравнивания временного ряда?

8. Что представляет собой автокорреляционная функция временного ряда?

9. Назовите этапы построения тренд-сезонных моделей временных рядов.

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

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

Задания для практики

Задача 1. 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 2. 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 3. 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 4. 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.

Глоссарий

Автокорреляция уровней ряда – корреляционная зависимость между уровнями временного ряда.

Аддитивная модель временного ряда – временной ряд представлен как сумма циклической, трендовой и случайной компонент.

Аналитическое выравнивание временного ряда – способ моделирования тенденции временного ряда посредством построения аналитической функции, характеризующей зависимость уровней ряда от времени.

Временной ряд – совокупность значений какого-либо показателя за несколько последовательных моментов времени.

Коррелограмма – график зависимости значений автокорреляционной функции временного ряда от величины лага.

Лаг – число периодов, по которым рассчитывается коэффициент корреляции временного ряда.

Мультипликативная модель временного ряда – временной ряд представлен как произведение циклической, трендовой и случайной компонент.

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

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

Тренд – это основная достаточно устойчивая тенденция во временном ряду, более или менее свободная от случайных колебаний.

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11844

http://tulpar.kpfu.ru/mod/resource/view.php?id=11845

2. Бородич С. А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006. – Гл. 12.

3. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 6,7.

Лекция 4(2)

Лекция 4(2). Модели стационарных и нестационарных временных рядов

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 выполнить практические задания.

•Для проверки усвоения темы имеется тест.

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).

Модели стационарных и нестационарных временных рядов, их идентификация

Набор случайных переменных X(t) называется стохастическим процессом. Стохастический процесс X_t называется стационарным в сильном смысле, если совместное распределение вероятностей всех переменных X_{t1}, X_{t2},…, X_{tn}, такое же, что и для переменных $X_{t1+\tau}$, $X_{t2+\tau}$,…, $X_{tn+\tau}$.

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 $x_1, x_2, ..., x_t$, т.е. конкретная реализация стационарного стохастического процесса X_t, также называется стационарным.

Стационарность означает отсутствие:			
тренда	систематиче с-ких изменений дисперсии	строго периодичны х колебаний	систематичес ки изменя-ющихся взаимозави-симостей между элементами временного ряда

Рис. 9.1. Признаки стационарности

Процесс называется нормальным, если совместное распределение $X_{t1}, X_{t2}, …, X_{tn}$ – это n-мерное нормальное распределение.

«Белым шумом» называется чисто случайный стационарный процесс, т е. ряд независимых, одинаково распределенных случайных величин a_i.

Модели стационарных временных рядов:

- модели авторегрессии порядка p (АР(p) – модели);
- модели скользящего среднего порядка q(СС(q)-модели);
- авторегрессионные модели со скользящими средними в остатках;
- (АРСС(p,q)-модели);
- простая и обобщенная модели авторегрессионных условно гетероскедастичных остатков.

АР(p) – модели:

P=1, марковский процесс: $\delta_t = \varepsilon_t - \alpha \cdot \varepsilon_{t-1}$;

P=2, процесс Юла: $\delta_t = \varepsilon_t - \alpha_1 \cdot \varepsilon_{t-1} - \alpha_2 \cdot \varepsilon_{t-2}$;

P≥3: $\delta_t = \varepsilon_t - \sum_{j=1}^{p} \alpha_j \varepsilon(t-j)$, где

δ - «белый шум»;

α- числовой коэффициент

ε_t - шок или импульс.

СС(q)-модели:

q=1: $\varepsilon_t = \delta_t - \theta \cdot \delta_{t-1}$;

q=2: $\varepsilon_t = \delta_t - \theta_1 \cdot \delta_{t-1} - \theta_2 \cdot \delta_{t-2}$;

q≥3: $\varepsilon_t = \delta_t - \theta_1 \cdot \delta_{t-1} - \theta_2 \cdot \delta_{t-2} - \ldots - \theta_q \cdot \delta_{t-q}$.

АРСС(p,q)- модель:

$\varepsilon_t = \alpha_1 \varepsilon_{t-1} + \ldots + \alpha_p \cdot \varepsilon_{t-p} + \delta_t - \theta_1 \delta_{t-1} - \ldots - \theta_q \delta_{t-q}$.

Модели нестационарных временных рядов:

- модель авторегрессии – проинтегрированного скользящего среднего (АРПСС(p,q,k) – модель);
- модели рядов, содержащих сезонную компоненту;
- регрессионные модели с распределенными лагами.

АРПСС(p,q,k) – модель (модель Бокса Дженкинса, ARIMA-модель) – это модель авторегрессии (левая часть) – проинтегрированного скользящего среднего (правая часть):

$$A_p(F\_,\alpha)x(t) = S^k(B_q(F\_,\theta)\delta(t)).$$

<u>Параметрические тесты на стационарность:</u>

1) 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:

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

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

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

2) 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.

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

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

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

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

$\bar{s}^2(n)$ - среднее значение временного ряда;

$\bar{y}$ - средняя дисперсия.

Проверка гипотезы о постоянстве дисперсии временного ряда в случае его разбиения на две части обычно осуществляется с использованием двухстороннего критерия Фишера.

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

$$F(\frac{1-p}{2}, v_1, v_2) \leq F \leq F(\frac{1+p}{2}, v_1, v_2) \Rightarrow H_0$$

$$H_0 : s_1^2 = s_2^2 = \sigma$$

При средних (от 40 до 100) и больших (более 100) объемах временного ряда для проверки гипотезы о постоянстве дисперсии временного ряда применяется стандартизованное нормальное распределение, H_0 – гипотеза о постоянстве дисперсии временного ряда.

$$\hat{O} = \frac{\frac{1}{2}\ln\frac{s_1^2}{s_2^2} + \frac{1}{2}(\frac{1}{v_1} - \frac{1}{v_2})}{\sqrt{\frac{1}{2}(\frac{1}{v_1} - \frac{1}{v_2})}},$$

$$|\hat{O}| < \hat{O}(p) \Rightarrow H_0$$

Модель авторегрессии–скользящего среднего (модель ARMA).

Использование моделей авторегресии основано на предположении о том, что текущее значение такого процесса может быть выражено в виде линейной комбинации предыдущих значений и случайной ошибки, обладающей свойствами «белого шума».

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:

$$\begin{aligned} & r_1 = a_1 + a_2 r_1 + ... + a_k r_{k-1}; \\ & r_2 = a_1 r_1 + a_2 + ... + a_k r_{k-2}; \\ & \ldots\ldots\ldots \\ & r_k = a_1 r_{k-1} + a_2 r_{k-2} + ... + a_k; \end{aligned}$$

$r_1, r_2, \ldots r_k$ – 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):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t,$$

$$r_1 = a_1 + a_2 r_1;$$

$$r_2 = a_1 r_1 + a_2.$$

$$a_1 = \frac{r_1(1 - r_2)}{1 - r_1^2};$$

$$a_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

В моделях скользящего среднего текущее значение стационарного процесса второго порядка y_t представлено в виде линейной комбинации текущего и прошедших значений значений ошибки ε, соответствующей «белому шуму».

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} ... - \beta_m \varepsilon_{t-m},$$

$$M(\varepsilon_t) = 0, D(\varepsilon_t) = \sigma_\varepsilon^2 = const,$$

$$\gamma i = M(\varepsilon_t, \varepsilon_{t-i}) = \begin{cases} \sigma_\varepsilon^2, i = 0 \\ 0, i \neq 0 \end{cases},$$

$$\rho_i^{(\varepsilon)} = 1, i = 0; \rho_i^{(\varepsilon)} = 0, i \neq 0.$$

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):

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

Модель авторегрессии - скользящего среднего АРСС(k,m) - (**A**uto **R**egressive-**M**oving **A**verage (ARMA (k,m))
Включает члены, описывающие авторегрессионные составляющие, и члены, моделирующие остаток в виде процесса скользящих средних. В ней, k – порядок авторегрессии, m – порядок скользящего среднего.

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_k y_{t-k} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} ... - \beta_m \varepsilon_{t-m}$$

Такая модель может интерпретироваться как линейная модель множественной регрессии, в которой в качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка – скользящие средние из элементов «белого шума». В экономических задачах чаще всего используют простейшую модель ARMA (1,1):

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

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).

Модель авторегрессии – принтегрированного скользящего среднего АРПСС(k,m,q) - (**A**uto **R**egressive **I**ntegrated **M**oving **A**verage (ARIMA (k,m,q)) или модель Бокса-Дженкинса. Экономические временные ряды чаще всего бывают нестационарными. Нестационарность проявляется в наличии неслучайной составляющей f(t). Если случайный остаток после вычитания f(t) представляет собой стационарный временной ряд, то исходный ряд называется нестационарным однородным. Для описания нестационарных однородных временных рядов применяется модель Бокса-Дженкинса (ARIMA –модель).

ARIMA –модель используется для описания временных рядов, обладающих следующими свойствами:

- ряд включает (аддитивно) составляющую f(t), имеющую вид алгебраического полинома;
- ряд после процедур последовательных разностей может быть описан моделью ARMA (k,m).

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.

У адекватной модели остатки должны быть похожими на «белый шум», то есть выборочные автокорреляции не должны существенно отличаться от нуля. Если модель адекватна исходным данным и ошибки являются «белым шумом», то распределение коэффициентов автокорреляции приближается к нормальному с нулевым математическим ожиданием и дисперсией 1/n.

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

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

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

<u>Критерии качества подгонки модели Бокса-Дженкинса:</u>

Критерий Акайка (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 (Swarz criterion):

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

АРПСС(p,q,k) – модель (модель Бокса Дженкинса, ARIMA-модель) – это модель авторегрессии (левая часть) – проинтегрированного скользящего среднего (правая часть):

$$A_p(F\_,\alpha)x(t) = S^k(B_q(F\_,\theta)\delta(t))$$

Вопросы для самоконтроля

1. Дайте определение стационарного временного ряда в узком и широком смысле слова.

2. Назовите виды моделей стационарных временных рядов.

3. Как выглядит модель, описывающая процесс Юла?

4. Охарактеризуйте поведение автокорреляционных функций при стационарности процесса AR(2).

5. Что такое процесс «случайного блуждания»?

6. Дайте определение авторегрессионного процесса скользящего среднего.

7. В чем отличие ARIMA-процессов от ARMA-процессов.

8. Какие методы оценивания параметров ARIMA-процессов наиболее предпочтительны?

9. Назовите критерии выбора наилучшей ARIMA-модели.

Задания для практики

Задача 1. Дана модель авторегрессии третьего порядка

$$y_t = 3y_{t-1} - 0{,}25y_{t-2} + 0{,}75y_{t-3} + \varepsilon_t.$$

Задание: построить характеристическое уравнение, найти его корни и установить, является ли указанный авторегрессионный процесс стационарным.

Задача 2. Для авторегрессии второго порядка

$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$ найдены выборочные значения автокорреляционной функции: $r(1) = 0{,}853$, $r(2) = 0{,}826$.

Задание: Оценить параметры авторегрессии, используя для этого уравнения Юла-Уолкера.

Задача 3. Задание: оценить параметры авторегрессии второго порядка $y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$ по следующим наблюдениям:

t	1	2	3	4	5	6	7	8	9	10	11	12	13
y_t	0,1	-2,5	-4	2,5	-0,2	-2,7	0,1	0,9	3,1	-0,5	1,8	0,5	1,9

Задача 4. Модель зависимости объемов продаж компании в среднем за месяц от расходов на рекламу была следующая (млн. руб):

$$\tilde{y}_t = -0{,}73 + 4{,}3x_t + 3{,}5x_{t-1} + 1{,}2x_{t-2} + 0{,}8x_{t-3}.$$

Задание: найти краткосрочный, долгосрочный мультипликатор и средний лаг.

Глоссарий

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

Авторегрессия – регрессия, учитывающая влияние предыдущих уровней на последующие.

Бокса Дженкинса модель – это модель авторегрессии (левая часть) – проинтегрированного скользящего среднего (правая часть), описывающая нестационарный однородный временной ряд.

Бокса-Пирса статистика – статистический критерий для обнаружения «белого шума» в остатках регрессии.

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

Модель авторегрессии – скользящего среднего – это линейная модель множественной регрессии, в которой в качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка – скользящие средние из элементов «белого шума».

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

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11855

2. Бородич С.А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006.- Гл. 12.

Лекция 5(1)

Лекция 5(1). Понятие о системах эконометрических уравнений

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.

Вопросы для изучения:

1. Понятие о системах уравнений. Системы независимых уравнений и системы взаимозависимых уравнений.
2. Структурная и приведенная формы модели.
3. Идентификация модели.

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

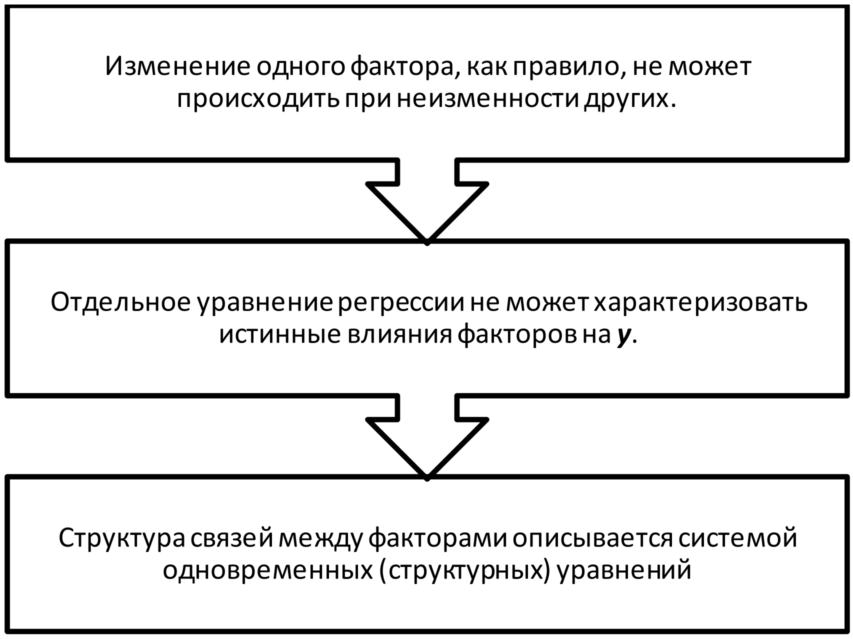

Рис. 10.1. Необходимость систем уравнений

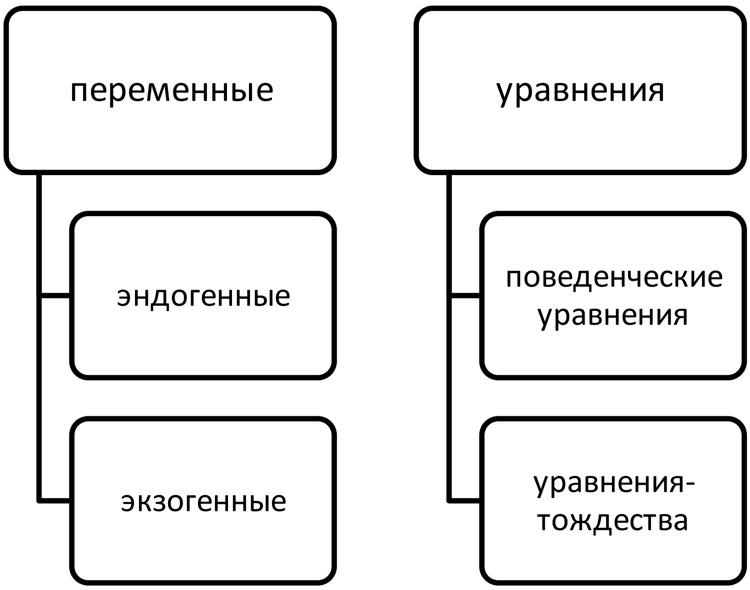

Рис. 10.2. Составляющие систем уравнений

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 (y_{t-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. 10.3. Виды систем уравнений

В системе независимых уравнений каждая зависимая переменная $\boldsymbol{y}$ рассматривается как функция одного и того же набора факторов $\boldsymbol{x}$:

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

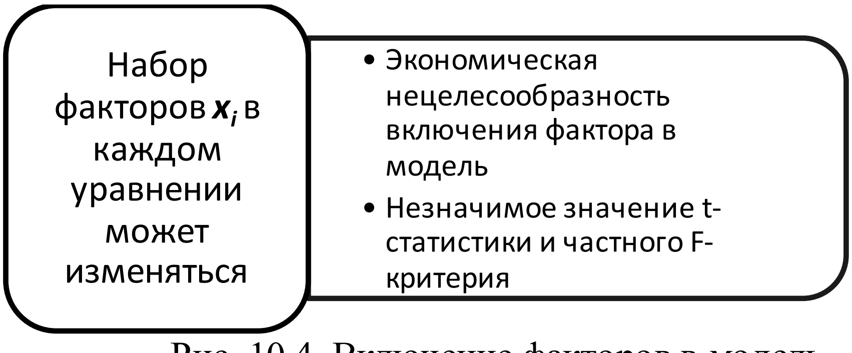

Рис. 10.4. Включение факторов в модель

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 ***y*** предшествующих уравнений наряду с набором собственно факторов ***x***:

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

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

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

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.

Все переменные в модели выражены в отклонениях от среднего уровня (x-$x_{ср}$; y-$y_{ср}$), поэтому свободный член в каждом уравнении отсутствует.

Структурная форма модели:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}\ x_2 + \varepsilon_{2.} \end{cases}$$

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

В каждое приведенное уравнение включаются все экзогенные переменные структурной модели.

Приведенные коэффициенты представляют собой нелинейные функции коэффициентов структурной модели. Приведенная форма модели:

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

Почему нужна приведенная форма модели?

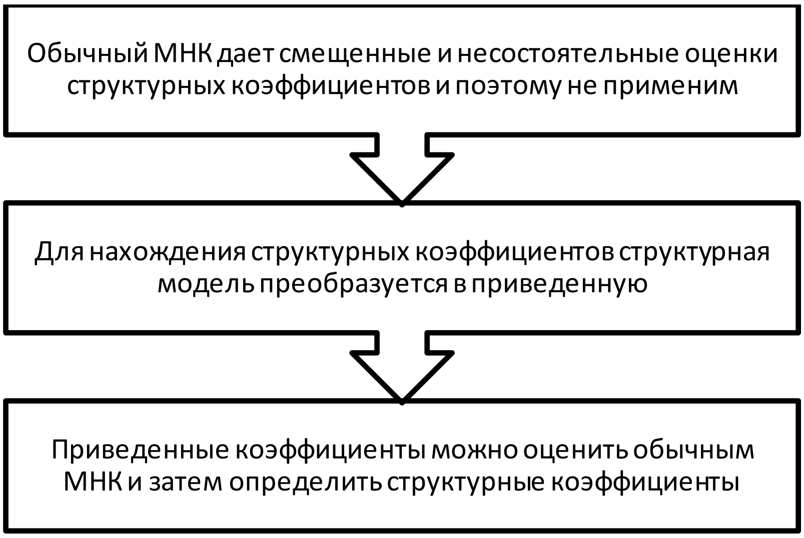

Рис. 10.5. МНК-оценки структурных коэффициентов

Идентификация модели

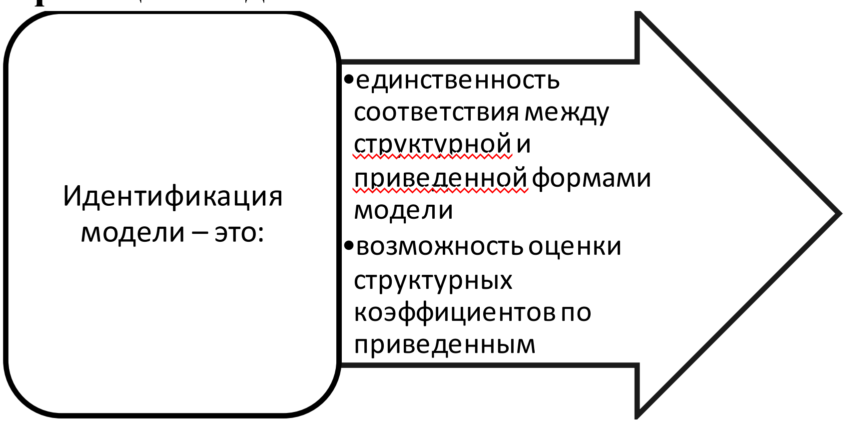

Рис. 10.6. Идентификация модели

Виды структурных моделей:

- идентифицируемые;
- неидентифицируемые;
- сверхидентифицируемые.

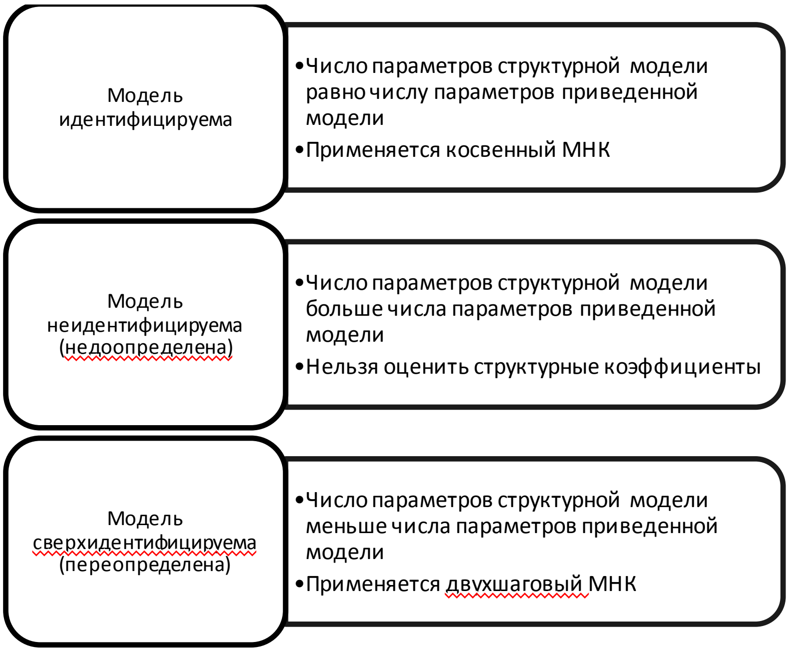

Рис. 10.7. Признаки идентификации

Необходимое условие идентификации: D+1=H - уравнение идентифицируемо; D+1<H - уравнение неидентифицируемо; D+1>H - уравнение сверхидентифицируемо. D - число экзогенных переменных, которые содержатся в системе, но не входят в данное уравнение; H - число эндогенных переменных в уравнении.

Достаточное условие идентификации: определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

Следует помнить, что на идентификацию проверяется каждое уравнение модели.

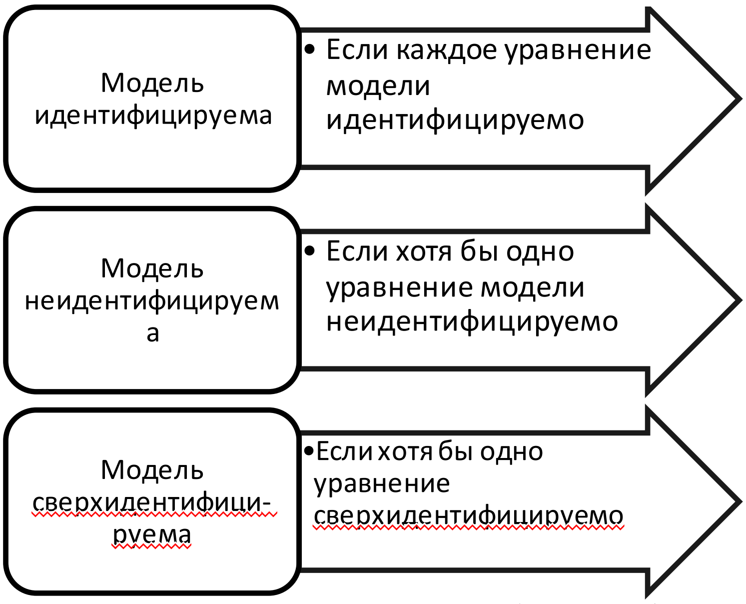

Рис. 10.8. Вывод об идентификации

Вопросы для самоконтроля

1. Перечислите возможные способы построения систем уравнений. Чем они отличаются друг от друга?

2. Как связаны между собой структурная и приведенная формы модели?

3. В чем суть косвенного МНК?

4. Когда модель считается идентифицируемой?

5. Какой метод оценки структурных коэффициентов и почему используется для точноидентифицируемой модели?

6. Когда выполняется необходимое условие идентификации ?

7. Какие переменные могут стоять в правой части структурной формы взаимозависимой системы?

8. Какие переменные могут стоять в правой части структурной формы системы независимых уравнений?

9. Когда можно использовать обычный МНК для оценки каждого из уравнений системы одновременных уравнений?

10. Когда выполняется достаточное условие идентификации?

Задания для практики

Задача 1. 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 2. 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 3.

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: Проверьте структурную форму модели на идентификацию.

Глоссарий

Достаточное условие идентификации – определитель матрицы, составленный из коэффициентов при переменных, отсутствующих в исследуемых уравнениях, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы. Для решения идентифицируемого уравнения применяется косвенный МНК, для решения сверхидентифицируемого – двухшаговый МНК.

Необходимое условие идентификации - выполнение счетного правила:D+1=H – уравнение идентифицируемо, D+1<H – уравнение неидентифицируемо, D+1>H – уравнение сверхидентифицируемо (H – число эндогенных переменных в уравнении, D – число предопределенных переменных, отсутствующих в уравнении, но присутствующих в системе).

Приведенная форма модели - система линейных функций эндогенных переменных от всех предопределенных переменных системы.

Система взаимосвязных (совместных) уравнений – одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других – в правую.

Система независимых уравнений – каждая зависимая переменная у рассматривается как функция одного и того же набора факторов х.

Система рекурсивных уравнений – зависимая переменная у одного уравнения выступает в виде фактора х в другом уравнении.

Экзогенные переменные – независимые переменные, которые определяются вне системы.

Эндогенные переменные – взамосвязные переменные которые определяются внутри модели.

Использованные информационные ресурсы

1. http://tulpar.kpfu.ru/mod/resource/view.php?id=11859

2. Бородич С.А. Эконометрика: Учебное пособие. -Мн.: Новое знание,2006,408 с. Глава 13.

3. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 9.

Лекция 5(2)

Лекция 5(2). Методы оценки систем одновременных уравнений

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.

Вопросы для изучения:

1. Косвенный, двухшаговый и трехшаговый МНК.

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

Косвенный, двухшаговый и трехшаговый МНК

Косвенный МНК применяется для оценивания идентифицируемых систем одновременных уравнений.

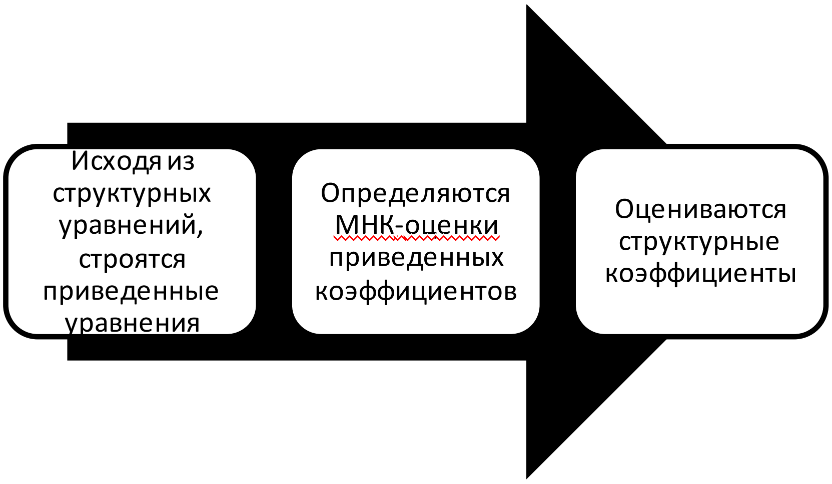

Рис. 11.1. Этапы косвенного МНК

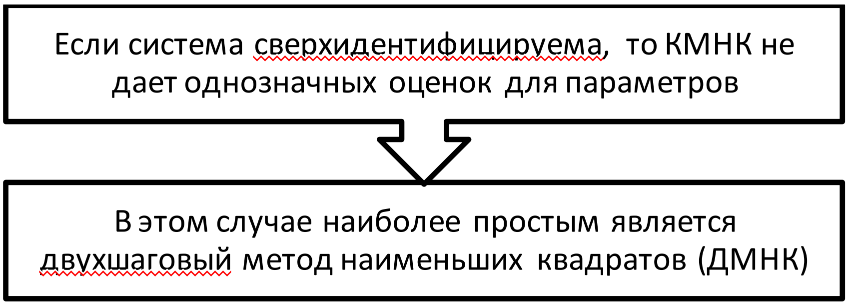

Рис. 11.2. Двухшаговый метод наименьших квадратов

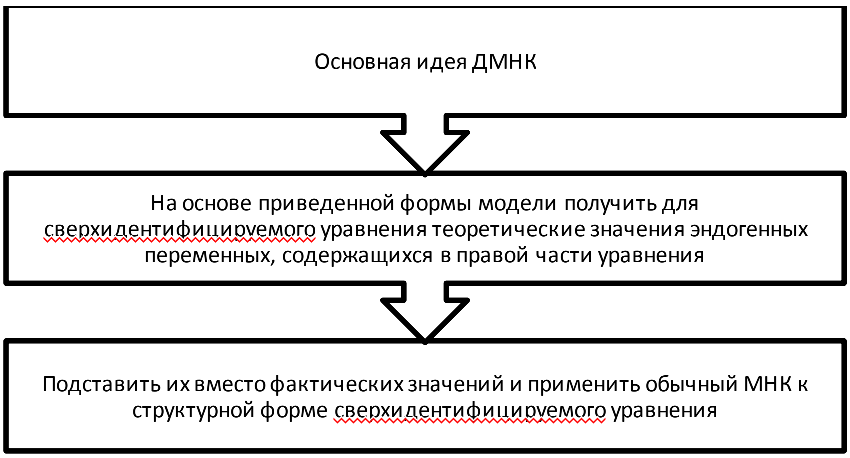

Рис. 11.3. Основная идея двухшагового метода наименьших квадратов

Таким образом, метод наименьших квадратов используется дважды:

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

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

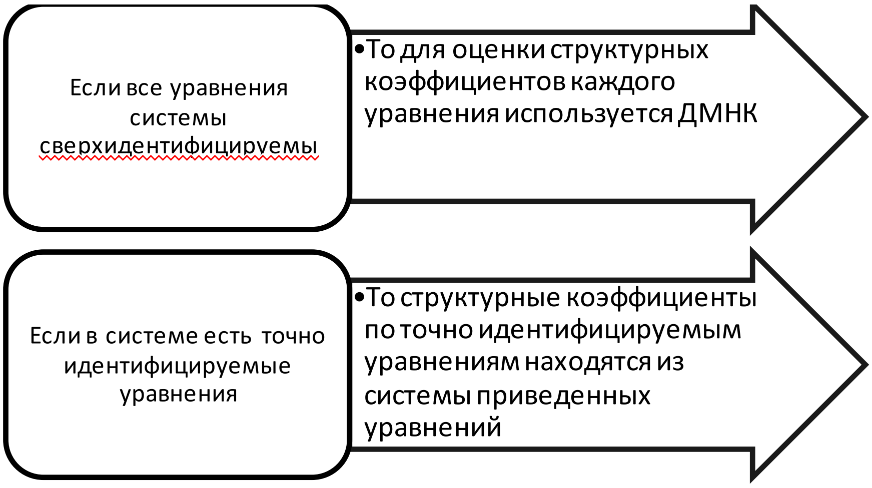

Рис. 11.4. Оценка структурных коэффициентов

Применим ДМНК к простейшей сверхидентифицируемой модели:

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

1 уравнение является сверхидентифицируемым: H=1 (y_1), D=1 (x_2) и D+1>H.

2 уравнение является точно идентифицируемым: H=2 (y_1, y_2), D=1 (x_1) и D+1=H.

Условные данные по пяти регионам

Регион	Y1	Y2	X1	X2
1	2	5	1	3
2	3	6	2	1
3	4	7	3	2
4	5	8	2	5
5	6	5	4	6
Средние	4	6,2	2,4	3,4

Приведенная форма модели составит:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2. \end{cases}$$

Используя отклонения от средних уровней, для первого уравнения приведенной формы модели система нормальных уравнений составит:

$$\begin{cases} \sum y_1x_1 = \delta_{11}\sum x_1^2 + \delta_{12}\sum x_1x_2, \\ \sum y_1x_2 = \delta_{11}\sum x_1x_2 + \delta_{12}\sum x_2^2. \end{cases}$$

$$\begin{cases} 6 = 5{,}2\delta_{11} + 4{,}2\delta_{12}, \\ 10 = 4{,}2\delta_{11} + 17{,}2\delta_{12}. \end{cases}$$

$$y_1 = 0{,}852x_1 + 0{,}373x_2 + u_1.$$

Используя отклонения от средних уровней, для второго уравнения приведенной формы модели система нормальных уравнений составит:

$$\begin{cases} \sum y_2x_1 = \delta_{21}\sum x_1^2 + \delta_{22}\sum x_1x_2, \\ \sum y_2x_2 = \delta_{21}\sum x_1x_2 + \delta_{22}\sum x_2^2. \end{cases}$$

$$\begin{cases} -0{,}4 = 5{,}2\delta_{21} + 4{,}2\delta_{22}, \\ -0{,}4 = 4{,}2\delta_{21} + 17{,}2\delta_{22}. \end{cases}$$

$$y_2 = -0{,}0728x_1 - 0{,}00557x_2 + u_2.$$

Таким образом, приведенная форма модели имеет вид:

$$\begin{cases} y_1 = 0{,}852x_1 + 0{,}373x_2 + u_1, \\ y_2 = -0{,}0728x_1 - 0{,}00557x_2 + u_2. \end{cases}$$

На основе второго уравнения данной системы можно найти теоретические значения (оценки) для эндогенной переменной y_2. Затем, используя сверхидентифицируемое структурное уравнение: $y_1=b_{12}(y_2+x_1)$, и заменив фактические значения y_2 их оценками, найдем значения новой переменной z: $\hat{y}_2 + x_1 = z$

Расчетные данные для второго шага ДМНК

X1	X2	Y2 (теорет)	Z	Y1	Y1Z	Z2
-1,4	-0,4	0,103	-1,297	-2	2,594	1,682
-0,4	-2,4	0,042	-0,358	-1	0,358	0,128
0,6	-1,4	-0,035	0,565	0	0	0,319
-0,4	1,6	0,020	-0,380	1	-0,380	0,144
1,6	2,6	-0,130	1,470	2	2,940	2,161
Сумма =0	0	0	0	0	5,512	4,434

Далее применим МНК к уравнению $y_1=b_{12}(y_2+x_1)$:

$$y_1 = b_{12}z,$$

$$\sum y_1z = b_{12}\sum z^2,$$

$$b_{12} = \frac{\sum y_1z}{\sum z_2} = \frac{5{,}512}{4{,}434} = 1{,}243.$$

Таким образом, первое сверхидентифицируемое структурное уравнение составит: $y_1 = 1{,}243(y_2 + x_1) + \varepsilon_1$.

Второе точно идентифицируемое структурное уравнение найдем из системы приведенных уравнений:

$$\begin{cases} y_1 = 0{,}852x_1 + 0{,}373x_2 + u_1, \\ y_2 = -0{,}0728x_1 - 0{,}00557x_2 + u_2. \end{cases}$$

С этой целью из второго уравнения приведенной формы модели следует исключить x_1, выразив его через первое уравнение и подставив во второе:

$$x_1 = \frac{y_1 - 0{,}373x_2}{0{,}852},$$

$$\hat{y}_2 = -0{,}0728\left(\frac{y_1 - 0{,}373x_2}{0{,}852}\right) - 0{,}00557x_2,$$

Таким образом, второе уравнение структурной формы модели:

$$\hat{y}_2 = -0{,}085y_1 + 0{,}026x_2 + \varepsilon_2$$

В целом рассматриваемая система одновременных уравнений составит:

$$\begin{cases} y_1 = 1{,}243(y_2 + x_1) + \varepsilon_1, \\ y_2 = -0{,}085y_1 + 0{,}026x_2 + \varepsilon_2. \end{cases}$$

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.

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

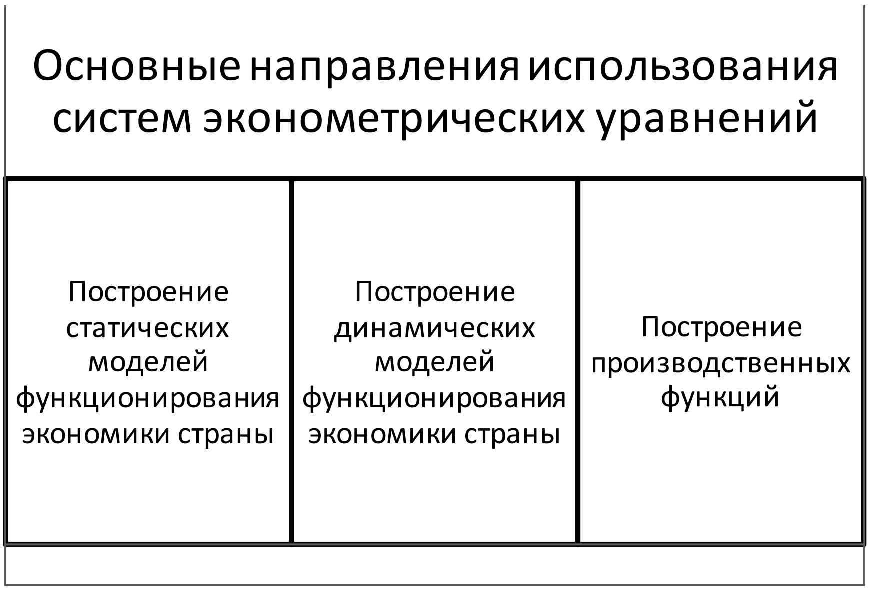

Рис. 11. 5. Основные направления использования систем эконометрических уравнений

Статическая модель Кейнса:

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

C – личное потребление в постоянных ценах;
y – национальный доход в постоянных ценах;
ε - случайная составляющая;
I – инвестиции в постоянных ценах.

В силу наличия тождества в модели структурный коэффициент b не может быть больше 1. Он характеризует предельную склонность к потреблению. Так, если b=0,65, то из каждой дополнительной 1 тыс. руб. дохода на потребление расходуется в среднем 650 руб. и 350 руб. инвестируется. Если b>1, то y< C+I, то есть на потребление расходуются не только доходы, но и сбережения. Структурный коэффициент b используется для расчета мультипликаторов. Инвестиционный мультипликатор потребления рассчитывается по формуле:

$$M_c = b/(1-b)$$
$$b = 0{,}65; M_c = 0{,}65/(1-0{,}65) = 1{,}857.$$

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

Инвестиционный мультипликатор национального дохода рассчитывается по формуле:

$$M_y = 1/(1-b)$$
$$b = 0{,}65; M_y = 1/(1-0{,}65) = 2{,}857.$$

Эта величина означает, что дополнительные инвестиции в размере 1 тыс. руб. на длительный срок приведут при прочих равных условиях к дополнительному доходу в 2,857 тыс. руб.

Динамическая модель Клейна:

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

C_t – функция потребления в период t; S_t – заработная плата в период t;P_t – прибыль в период t; P_{t-1} – прибыль в период t-1; Rt – общий доход в период t; R_{t-1} – общий доход в предыдущий период;t – время;Tt – чистые трансферты в пользу администрации в период t; I_t – 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:

$$\begin{cases} C_t = d_1T + d_2G + d_3t + d_4P_{t-1} + d_5R_{t-1} + u_1, \\ I_t = d_6T + d_7G + d_8t + d_9P_{t-1} + d_{10}R_{t-1} + u_2, \\ S_t = d_{11}T + d_{12}G + d_{13}t + d_{14}P_{t-1} + d_{15}R_{t-1} + u_3, \\ R_t = d_{16}T + d_{17}G + d_{18}t + d_{19}P_{t-1} + d_{20}R_{t-1} + u_4, \\ P_t = d_{21}T + d_{22}G + d_{23}t + d_{24}P_{t-1} + d_{25}R_{t-1} + u_5, \end{cases}$$

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 d_1, d_6, d_{11}, d_{16}, d_{21} – мультипликаторы чистых трансфертов в пользу администрации относительно личного потребления d_1, инвестиций d_6, заработной платы d_{11}, дохода d_{16} и прибыли d_{21}. Соответственно коэффициенты d_2, d_7, d_{12}, d_{17}, d_{22} являются мультипликаторами правительственных расходов относительно соответствующих эндогенных переменных.

Динамическая модель Кейнса:

$$\begin{cases} C_t = a + b_1Y_t + b_2Y_{t-1} + \varepsilon_1, \\ Y_t = C_t + G_t + I_t + L_t, \\ P_t = Y_t + Z_t. \end{cases}$$

Yt – имеющийся в распоряжении доход в период времени t; C_t – частное потребление в период времени t; P_t – валовой национальный продукт в период времени t; Y_{t-1} – доход предыдущего года; Gt - общественное потребление; I_t – валовые капиталовложения; L_t – 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:

$$\begin{cases} Q^d = a_0 + a_1P + \varepsilon_1, \\ Q^s = b_0 + b_1P + \varepsilon_2, \\ Q^d = Q^s. \end{cases}$$

Qd – спрашиваемое количество благ (объем спроса);
Q_s – предлагаемое количество благ (объем предложения).
В этой системе три эндогенные переменные – Qd, Q_s и P. При этом если Qd и Q_s 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).

Вопросы для самоконтроля

1. Для оценки каких систем возможно применять обычный МНК?
2. В чем суть косвенного МНК?
3. Всегда ли можно применить косвенный МНК?
4. В чем суть двухшагового МНК и когда он применяется?
5. Что представляют собой мультипликаторные модели кейнсианского типа?
6. Приведите пример динамической модели экономики.

Задания для практики

Задача 1. Имеется следующая модель:

$$y_1 = \alpha_1 + a_{11}x_1 + a_{12}x_2 + b_{12}y_2 + \varepsilon_1,$$
$$y_2 = \alpha_2 + a_{22}x_2 + a_{23}x_3 + b_{21}y_1 + \varepsilon_2,$$
$$y_3 = \alpha_3 + a_{31}x_1 a_{33}x_3 + \varepsilon_3.$$

Приведенная форма этой модели имеет вид:

$$y_1 = 6 + 8x_1 + 10x_2 + 4x_3 + \upsilon_1,$$
$$y_2 = 16 - 12x_1 - 70x_2 + 8x_3 + \upsilon_2,$$
$$y_3 = 10 - 5x_1 - 22x_2 + 5x_3 + \upsilon_3.$$

Задание: определить все возможные структурные коэффициенты на основе приведенной формы модели.

Задача 2. Имеется следующая гипотетическая структурная модель:

$$y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + \varepsilon_1,$$
$$y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2 + \varepsilon_2,$$
$$y_3 = b_{32}y_2 + a_{31}x_1 + a_{33}x_3 + \varepsilon_3.$$

Приведенная форма исходной модели имеет вид:

$$\tilde{y}_1 = 3x_1 - 6x_2 + 2x_3,$$
$$\tilde{y}_2 = 2x_1 + 4x_2 + 10x_3,$$
$$\tilde{y}_3 = -5x_1 + 6x_2 + 5x_3.$$

Задание:

1) проверить структурную форму модели на идентифицируемость;
2) определить структурные коэффициенты модели.

Задача 3. Рассматривается следующая модель:

$$y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1,$$
$$y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2 + \varepsilon_2,$$
$$y_3 = b_{31}y_1 + b_{32}y_2 + a_{33}x_3 + \varepsilon_3.$$

Приведенная форма этой модели имеет вид:

$$\tilde{y}_1 = 2x_1 + 3x_2 + 4x_3,$$
$$\tilde{y}_2 = x_1 + 4x_2 + 8x_3,$$
$$\tilde{y}_3 = 2x_1 + 4x_2 + x_3.$$

Известно, что второе и третье уравнение точно идентифицируемы.

Задание: определить оценки коэффициентов структурной формы этих уравнений косвенным методом наименьших квадратов.

Задача 4. Задание: оценить двухшаговым методом наименьших квадратов структурные коэффициенты модели:

$$y_1 = \alpha_1 + b_{12}y_2 + a_{11}x_1 + \varepsilon_1,$$
$$y_2 = \alpha_2 + b_{21}y_1 + a_{22}x_2 + \varepsilon_2$$

при наличии следующих данных:

t	y_1	y_2	x_1	x_2
1	2	4	1	0
2	3	4	1	1
3	3	3	1	1
4	4	2	0	2
5	5	2	0	2

Глоссарий

Двухшаговый МНК: 1) составляется приведенная форма модели и определяются численные значения параметров каждого ее уравнения обычным МНК, 2) выявляются эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяются двухшаговым МНК, и находятся расчетные значения по соответствующим уравнениям приведенной формы модели, 3) обычным МНК определяются параметры структурного уравнения, используя в качестве исходных данных фактические значения предопределенных переменных и расчетные значения эндогенных переменных, стоящих в правой части структурного уравнения.

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

Косвенный МНК: 1) составляется приведенная форма модели и определяются численные значения параметров каждого ее уравнения обычным МНК, 2) путем алгебраических преобразований делается переход от приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

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

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

Использованные информационные ресурсы

1. Бородич С.А. Эконометрика: учебное пособие. -Мн.: Новое знание, 2006.- Гл. 13.

2. Эконометрика: учеб. /под ред. И. И. Елисеевой. -М.: Проспект, 2010.- Гл. 9.

Глоссарий по учебной дисциплине

ANCOVA-модель – это регрессионная модель, в которой объясняющие переменные носят как количественный, так и качественный характер.

Автокорреляция остатков регрессии – зависимость случайных отклонений ε_i и ε_j друг от друга для $i \neq j$.

Автокорреляция уровней ряда – корреляционная зависимость между уровнями временного ряда.

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

Авторегрессия – регрессия, учитывающая влияние предыдущих уровней на последующие.

Аддитивная модель временного ряда – временной ряд представлен как сумма циклической, трендовой и случайной компонент.

Аналитическое выравнивание временного ряда – способ моделирования тенденции временного ряда посредством построения аналитической функции, характеризующей зависимость уровней ряда от времени.

Базовое значение качественной переменной имеет цифровую метку ноль: D=0.

Бета-коэффициент показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение результативного признака при изменении факторного признака на величину своего среднеквадратического отклонения.

Бокса Дженкинса модель – это модель авторегрессии (левая часть) – проинтегрированного скользящего среднего (правая часть), описывающая нестационарный однородный временной ряд.

Бокса-Кокса подход - формализованная процедура подбора линеаризующего преобразования.

Бокса-Пирса статистика – статистический критерий для обнаружения «белого шума» в остатках регрессии.

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

Внутренне линейная нелинейная модель с помощью преобразований может быть приведена к линейному виду.

Внутренне нелинейная модель не может быть сведена к линейной функции.

Временной ряд – совокупность значений какого-либо показателя за несколько последовательных моментов времени.

Выборка – это часть генеральной совокупности.

Голдфелда Квандта тест – один из наиболее распространенных способов тестирования остатков регрессии на гетероскедастичность

Гомоскедастичность остатков регрессии – постоянство дисперсии случайных отклонений εi.

Дарбина-Уотсона тест – один из наиболее распространенных способов тестирования остатков регрессии на автокорреляцию

Двухшаговый МНК: 1) составляется приведенная форма модели и определяются численные значения параметров каждого ее уравнения обычным МНК, 2) выявляются эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяются двухшаговым МНК, и находятся расчетные значения по соответствующим уравнениям приведенной формы модели, 3) обычным МНК определяются параметры структурного уравнения, используя в качестве исходных данных фактические значения предопределенных переменных и расчетные значения эндогенных переменных, стоящих в правой части структурного уравнения.

Дискретная зависимая переменная – это переменная, которая принимает несколько альтернативных значений.

Дифференциальный коэффициент свободного члена – это коэффициент перед фиктивной переменной в регрессионной модели. Он показывает, на какую величину отличается свободный коэффициент ***a*** при значении D=1, от свободного коэффициента ***a*** при D=0.

Дифференциальный угловой коэффициент – это коэффициент перед произведением фиктивной переменной и независимой переменной в регрессионной модели. Он показывает, на какую величину отличается коэффициент регрессии ***b*** при значении D=1, от коэффициента регрессии ***b*** при D=0.

Достаточное условие идентификации – определитель матрицы, составленный из коэффициентов при переменных, отсутствующих в исследуемых уравнениях, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы. Для решения идентифицируемого уравнения применяется косвенный МНК, для решения сверхидентифицируемого – двухшаговый МНК.

Идентификация модели – проведение статистического анализа модели и оценивания качества ее параметров.

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

Индекс корреляции – показатель тесноты статистической взаимосвязи, выраженной в нелинейной форме.

Индекс множественной корреляции оценивает тесноту совместного влияния факторов на результативный признак Y.

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

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

Коррелограмма – график зависимости значений автокорреляционной функции временного ряда от величины лага.

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

Косвенный МНК: 1) составляется приведенная форма модели и определяются численные значения параметров каждого ее уравнения обычным МНК, 2) путем алгебраических преобразований делается переход от приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

Коэффициент детерминации – это показатель, который определяет долю разброса зависимой переменной Y, объясняемую регрессией Y на X.

Коэффициент эластичности – показатель, который измеряет, на сколько процентов изменится результативный признак Y, если факторный признак изменится на 1 %.

Коэффициент эластичности показывает, на сколько процентов изменяется результативный признак Y при изменении факторного признака X на один процент.

Лаг – число периодов, по которым рассчитывается коэффициент корреляции временного ряда.

Линеаризация – приведение нелинейных моделей регрессии к линейному виду путем замены, логарифмирования переменных или комбинированным способом с целью применения МНК.

Линейный коэффициент парной корреляции – это показатель тесноты статистической взаимосвязи между переменными Y и X.

Ловушка фиктивной переменной – это состояние совершенной мультиколлинеарности в силу строгой линейной зависимости между переменными D_1 и D_2, при котором коэффициенты уравнения регрессии однозначно определены быть не могут.

Логит-модель основана на использовании функции логистического распределения.

Метод максимального правдоподобия – один из способов оценивания параметров регрессии, в частности, в моделях с дискретной зависимой переменной.

Метод наименьших квадратов – один из распространенных способов оценивания параметров регрессии.

Множественная корреляция – это зависимость между результативным признаком и двумя и более факторными признаками.

Множественная регрессия представляет собой модель, где теоретическое (среднее) значение зависимой переменной Y рассматривается как функция нескольких независимых переменных $X_1, X_2,...X_m$.

Модели бинарного выбора содержат зависимую переменную, которая принимает лишь два альтернативных значения, обозначаемых цифровыми метками: 0 и 1.

Модели множественного выбора содержат зависимую переменную, которая принимает несколько упорядоченных или неупорядоченных альтернативных значений.

Модель авторегрессии – скользящего среднего – это линейная модель множественной регрессии, в которой в качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка – скользящие средние из элементов «белого шума».

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

Мультиколлинеарность - это линейная взаимосвязь двух или нескольких объясняющих переменных ($x_1, x_2, \dots x_m$).

Мультипликативная модель временного ряда – временной ряд представлен как произведение циклической, трендовой, случайной компонент.

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

Необходимое условие идентификации - выполнение счетного правила: $D+1=H$ – уравнение идентифицируемо, $D+1 < H$ – уравнение неидентифицируемо, $D+1 > H$ – уравнение сверхидентифицируемо (H – число эндогенных переменных в уравнении, D – число предопределенных переменных, отсутствующих в уравнении, но присутствующих в системе).

Однородные статистические данные – это совокупность данных, зарегистрированных при одних и тех же условиях.

Параметризация модели – выражение в математической форме взаимосвязи между переменными модели, формулирование исходных предпосылок и ограничений модели.

Парная регрессия представляет собой модель, где теоретическое (среднее) значение зависимой переменной Y рассматривается как функция одной независимой переменной X.

Парный коэффициент регрессии показывает, на какую величину в среднем изменится результативный признак Y, если переменную X увеличить на единицу измерения.

Предопределенные переменные – это экзогенные переменные и лаговые эндогенные переменные.

Приведенная форма модели - система линейных функций эндогенных переменных от всех предопределенных переменных системы.

Пробит-модель основана на использовании функции стандартного нормального распределения.

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

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

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

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

Система взаимосвязаhных уравнений – одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других – в правую.

Система независимых уравнений – каждая зависимая переменная у рассматривается как функция одного и того же набора факторов х.

Система рекурсивных уравнений – зависимая переменная у одного уравнения выступает в виде фактора х в другом уравнении.

Спецификация модели - формулирование вида модели, исходя из соответствующей теории связи между переменными.

Тренд – это основная достаточно устойчивая тенденция во временном ряду, более или менее свободная от случайных колебаний.

Фиктивные переменные - качественные переменные, преобразованные в количественные с помощью цифровых меток.

Цель регрессионного анализа – оценка функциональной зависимости между независимыми переменными X и условным математическим ожиданием зависимой переменной Y.

Цель эконометрики – эмпирический (практический) вывод экономических законов.

Частное уравнение регрессии характеризует изолированное влияние фактора X_j на результат.

Частный коэффициент эластичности показывает, на сколько % изменяется в среднем результативный признак Y при изменении фактора Xj на 1%.

Чоу тест – это статистический тест, определяющий целесообразность использования фиктивной переменной.

Экзогенные переменные – независимые переменные, которые определяются вне системы.

Экзогенные переменные - это внешние для модели переменные, управляемые из вне, влияющие на эндогенные переменные, но не зависящие от них.

Эконометрика – это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов, которые раскрыты и обоснованы экономической теорией.

Эндогенные переменные – взаuмосвязhные переменные, которые определяются внутри модели.

Эндогенные переменные - это внутренние, формируемые в модели переменные, зависимые от предопределенных переменных.