Московская финансово-промышленная академия

Кафедра Математических методов принятия решений

Копнова Е.Д.

Интернет-курс по дисциплине «Эконометрика»

Москва, 2010

Содержание

Аннотация к дисциплине

Цель преподавания курса:

Дать студентам научное представление о методах и моделях, позволяющих получать количественные выражения закономерностей экономической теории на базе статистики с использованием математико-статистического инструментария.

Необходимые предметы, предваряющие курс эконометрики:

· экономическая теория, дающая представление о направлениях развития экономики,

· статистика, в которой сформулированы общие методы и принципы определения количественных характеристик массовых процессов и явлений;

· высшая математика, в т.ч.

· линейной алгебры для проведения расчетов над матрицами;

· математического анализа, обучающего приемам интегрирования и дифференцирования;

· теории вероятностей для оперирования со случайными величинами;

· математической статистики, определяющей методы обработки выборочных данных, и распространения результатов их анализа на исследуемые явления и процессы.

Задачи курса:

· Научиться строить эконометрические модели.

· Научиться оптимизировать эконометрические модели.

· Научиться содержательно интерпретировать формальные результаты эконометрического моделирования.

· Научиться использовать эконометрический инструментарий пакетов прикладных программ (Excel, STATISTICA, SPSS, EViews и др.).

Приобретаемые профессиональные компетенции:

· Владение методами количественной оценки взаимосвязей между экономическими показателями на основе статистической информации об изучаемых объектах.

· Владение эконометрическими методами разработки прогноза развития предприятий на основе анализа их опыта работы.

· Владение инструментарием, позволяющим оперативно принимать обоснованные решения в экономике.

Материалы по темам

Тема 1. Эконометрическое моделирование

Основные задачи эконометрики. Классификация переменных в эконометрических моделях. Классификация эконометрических моделей. Корреляционная зависимость. Спецификация регрессионной модели.

Цель изучения:

Введение в предмет и методы эконометрики.

Задачи изучения:

· определить цели и задачи эконометрики, ее место в сфере социально-экономических исследований;

· определить базовые понятия эконометрики, необходимые для изучения основных тем курса.

Теоретический материал

Эконометрика – наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов при помощи методов статистического анализа.

Слово «эконометрика» введено в 1926 году норвежским экономистом и статистиком, лауреатом Нобелевской премии Рагнаром Фришем. Это слово означает – измерения в экономике.

Цель эконометрического анализа:

Разработка *эконометрических моделей*, позволяющих решить следующие основные задачи:

· Проверка экономических теорий

· Прогнозирование экономического развития

· Выработка рекомендаций по экономической политике.

Современное экономическое образование на западе держится на трех китах: макроэкономике, микроэкономике и эконометрике.

Связь с другими науками:

Эконометрика базируется на 3-х дисциплинах:

· Экономическая теория.

· Статистика - сбор и обработка информации (среднее значение, дисперсия, показатели корреляции).

· Математика – использование аппарата математического анализа, матричной алгебры, теории вероятности.

Этапы эконометрического моделирования:

· Определение цели исследования.

· **Качественный анализ зависимостей между экономическими показателями.**

· Спецификация модели – выбор переменных и связей между ними.

· **Сбор исходных данных, их анализ.**

· **Идентификация модели – статистический анализ модели (оценка параметров).**

· Оценка качества модели.

· Интерпретация модели и ее использование для прогнозирования.

Классификация переменных:

· По количеству переменных для каждого объекта:

- Одномерные.
- Двумерные.
- Множественные.

· По типу измерения:

- В номинальной шкале.
- В порядковой шкале.
- В интервальной шкале.

· По упорядоченности во времени:

- Пространственные данные - в данный момент по разным объектам.
- Временной ряд - наблюдения во времени за одним объектом.
- Панельные данные - сведения по разным объектам за несколько периодов.

· По источнику информации:

- Объясняющие.
- Объясняемые.
- Случайные возмущения.

· По соотношению в связи:

- Зависимые - результативный признак.
- Независимые – факторный признак.

· По отношению к модели:

- *Эндогенные* - значения которых объясняются в рамках модели.
- *Экзогенные* - значения которых являются для модели внешними.
- *Предопределенные* - экзогенные переменные и лаговые значения эндогенных переменных.

Корреляционная зависимость и эконометрическая модель.

Удобным графическим средством анализа данных является *диаграмма рассеяния.*

Вытянутость облака точек позволяет сделать предположение о существовании тенденция линейной связи между x и y

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

Выберем $(\bar{x}, \bar{y})$,. Если (x_i, y_i) правее вертикальной секущей, то $x_i - \bar{x} > 0$, левее, то <0. Если (x_i, y_i) выше горизонтальной секущей, то $y_i - \bar{y} > 0$, ниже, то <0.

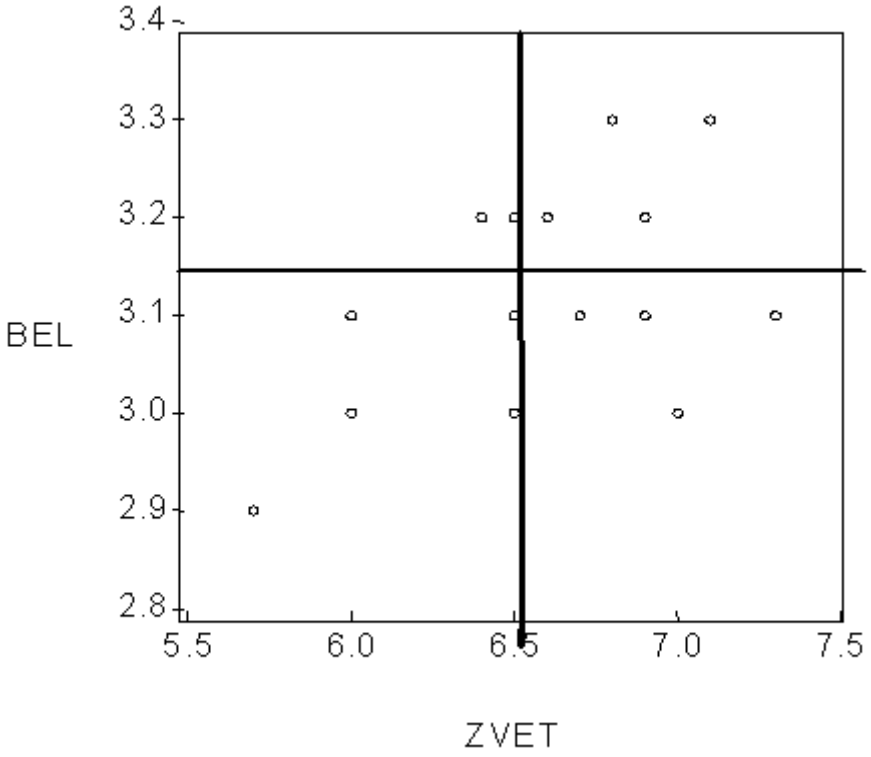
m_{++} - число точек, для которых $x_i - x > 0$ и $y_i - y > 0$;

m_{+-} - $x_i - x > 0$ и $y_i - y < 0$; m_{-+} — $x_i - x < 0$ и $y_i - y > 0$; m_{--} - $x_i - x < 0$ и $y_i - y < 0$.

$m_{++} = 4$, $m_{+-} = 4$, $m_{-+} = 3$, $m_{--} = 6$, $m_{++} + m_{--} = 10$, а $m_{+-} + m_{-+} = 7$.

Количество точек с совпадающими знаками отклонений от средних значений 10/17=0.59, т. е. около 59% общего числа точек - положительный угловой коэффициент.

Если бы большинство составляли точки с противоположными знаками - отрицательный угловой коэффициент.



Последняя ситуация наблюдается при рассмотрении зависимости спроса на товар от его цены.

Используют коэффициент корреляции

$$r_{xy} = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)}\sqrt{\mathrm{Var}(y)}}.$$

Величина $\mathrm{Cov}(x,y)$, определяется соотношением и называется ковариацией переменных x и y.

$$\mathrm{Cov}(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

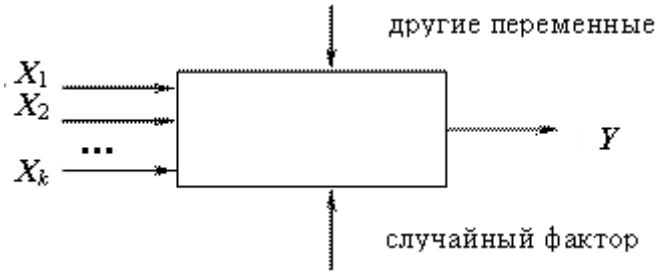
$$\mathrm{Cov}(x,x) = \mathrm{Var}(x), \quad \mathrm{Cov}(y,y) = \mathrm{Var}(y).$$

Если тенденция выражена ясно, то r_{xy} по абсолютной величине близки к единице. Если наличие линейной тенденции связи обнаруживается с трудом, то тогда значения r_{xy} близки к нулю.

Значение коэффициента	0,1 - 0,3	0,3 - 0,5	0,5 - 0,7	0,7 - 0,9	0,9 - 0,999
Характеристика связи	слабая	умеренная	заметная	высокая	весьма высокая

Эконометрическая модель - это математическая модель, которая отражает влияние факторов на результат функционирования экономической системы.

В модели принято выделять существенные и несущественные факторы. Существенные факторы $X_1, X_2 \ldots X_k$ формируют среднее значение результата $\bar{y}$. Функция $\bar{y} = \bar{y}(x_1, x_2 \ldots x_k)$ характеризует влияние каждого набора значений факторов на среднее значения результата. Такого рода зависимость называется *корреляционной зависимостью.*



Случайная компонента ε обеспечивает разброс результативного значения около среднего.

Эконометрическая модель представляет собой выражение вида:

$$y = \bar{y}(x_1, x_2 \ldots x_k, \varepsilon).$$

Классификация эконометрических моделей

· По виду данных:
- Регрессионные модели.
- Модели временных рядов.

· По числу факторов:
- Парная регрессия.
- Множественная регрессия.

· По виду зависимости:
- Линейные модели.
- Нелинейные модели.

· По использованию предыдущих шагов:
- Регрессионная модель - по данным текущего шага.
- Авторегрессионная модель - по данным текущего и предыдущих шагов.
- Модель скользящей средней – использует ошибки текущего и предыдущих шагов.

· По числу уравнений:
- Регрессионные модели с одним уравнением.
- Системы одновременных уравнений.

Примеры эконометрических моделей.

Конъюнктурная модель Клейна.

Наиболее часто в экономических исследованиях применяется конъюнктурная модель Клейна, разработанная в начале 50-х гг. XX в. для США.

$$\begin{cases} C_t = c_0 + c_1 P_t + c_2 P_{t-1} + c_3 (W_t^P + W_t^G) + \varepsilon_t^1, M\varepsilon_t^1 = 0, D\varepsilon_t^1 = \sigma_1^2, \\ I_t = i_0 + i_1 P_t + i_2 P_{t-1} + i_3 K_{t-1} + \varepsilon_t^1, M\varepsilon_t^2 = 0, D\varepsilon_t^2 = \sigma_2^2, \\ W_t^P = w_0 + w_1 Y_t + w_2 Y_{t-1} + w_3 t + \varepsilon_t^3, M\varepsilon_t^3 = 0, D\varepsilon_t^3 = \sigma_3^2, \\ Y_t = C_t + I_t + G_t, \\ P_t = Y_t - T_t - W_t^P, \\ K_t = K_{t-1} + I_t, t = 1, \ldots, n, \end{cases}$$

где C_t – потребление;

I_t – чистые инвестиции;

W_t^P – заработная плата в частном секторе;

W_t^G – заработная плата в государственном секторе;

Y_t – валовой внутренний продукт (без чистого экспорта и прироста запасов);

P_t – общая прибыль;

K_t – капитал;

G_t – государственные расходы;

T_t – общий сбор налогов.

В модели девять переменных и шесть уравнений. В число *эндогенных* переменных входят C_t, I_t, W_t^P, Y_t, P_t, K_t. Три из них Y_t P_t K_t являются лаговыми эндогенными, поскольку в текущий момент t принимают участие прошлые значения этих переменных $Y_{t-1}, P_{t-1} K_{t-1}$

Экзогенными переменными являются W_t^G, G_t, T_t, t. Они вместе с прошлыми значениями лаговых эндогенных переменных $Y_{t-1}, P_{t-1} K_{t-1}$ образуют набор $W_t^G, G_t, T_t, t. Y_{t-1}, P_{t-1} K_{t-1}$ *предопределенных* переменных.

Первые три уравнения содержат случайные составляющие $\varepsilon_1^1, \varepsilon_1^2, \varepsilon_1^3$. Последние три таких составляющих не содержат, поэтому являются балансовыми.

Модель Клейна, идентифицированная по данным Канады за .955-1975 гг., имеет следующий вид* (все стоимостные показатели указаны в млрд. долл. в ценах 1975 г.):

$$\begin{cases} C_t = 1,407 + 0,694 P_t + 0,1 P_{t-1} + 0,855 (W_t^P + W_t^G) + \varepsilon_t^1, \sigma_1 = 1,7, \\ I_t = -2,215 + 0,433 P_t + 0,947 P_{t-1} - 0,34 K_{t-1} + \varepsilon_t^2, \sigma_2 = 3,85, \\ W_t^P = 11,624 + 0,779 Y_t - 0,159 Y_{t-1} + 0,698_t + \varepsilon_t^3, \sigma_3 = 1,87, \\ Y_t = C_t + I_t + G_t, \\ P_t = Y_t - T_t - W_t^P, \\ K_t = K_{t-1} + I_t. \end{cases}$$

Увеличение текущей прибыли на 1 млрд. долл. приводит к среднему увеличению потребления на 694 млн. долл., а такое же увеличение фонда заработной платы в частном и государственном секторах — к среднему росту потребления на 855 млн. долл.

На рост инвестиций наибольшее влияние оказывает прибыль прошлого года, а на рост фонда заработной платы в частном секторе — ВВП текущего года, кроме того, имеется тенденция среднегодового роста этого фонда на 698 млн. долл.

Рынок квартир в Москве.

Данные для этого исследования собраны студентами РЭШ Российской экономической школы) в 1996 гг. После проведенного анализа была выбрана логарифмическая форма модели, как более соответствующая данным:

$$LOGPRICE = \beta_0 + \beta_1 LOGLIVSP + \beta_2 LOGPLAN + \beta_3 LOGKITSP + \beta_4 LOGDIST + \beta_5 FLOOR + \beta_6 BRICK + \beta_7 BAL + \beta_8 LIFT + \beta_9 R1 + \beta_{10} R2 + \beta_{11} R3 + \beta_{12} R4 + \varepsilon$$

Здесь *LOGPRICE* — логарифм цены квартиры (в долл. США),
LOGLIVSP — логарифм жилой площади (в кв.м),
LOGPLAN — логарифм площади нежилых помещении (в кв.м),
LOGKITSP — логарифм площади кухни (в кв.м),
LOGDIST — логарифм расстояния от центра Москвы (в км).

Включены также бинарные, "фиктивные" переменны, принимающие значения 0 или 1:

FLOOR — принимает значение 1, если квартира расположена на первом или на последнем этаже,

BRICK — принимает значение 1, если квартира находится в кирпичном доме,

BAL — принимает значение 1, если в квартире есть балкон,

LIFT — принимает значение 1, если в доме есть лифт,

R1 — принимает значение 1 для однокомнатных квартир и 0 для всех остальных,

R2, R3, R4 — аналогичные переменные для двух-, трех- и четырехкомнатных квартир.

Результаты оценивания уравнения (1.5) для 464 наблюдений, относящихся к 1996 г., приведены в таблице 1.

Таблица 1

Переменная	Коэффициент	Стандартная ошибка	t-статистика	P-значение
CONST	7.106	0.290	24.5	0.0000
LOGLIVSP	0.670	0.069	9.65	0.0000
LOGPLAN	0.431	0.049	8.71	0.0000
LOGKITSP	0.147	0.060	2.45	0.0148
LOGDIST	-0.114	0.016	-7.11	0.0000
BRICK	0.134	0.024	5.67	0.0000
FLOOR	-0.0686	0.021	-3.21	0.0014

LIFT	0.114	0.024	4.79	0.0000
BAL	0.042	0.020	2.08	0.0385
R1	0.214	0.109	1.957	0.0510
R2	0.140	0.080	1.75	0.0809
R3	0.164	0.060	2.74	0.0065
R4	0.169	0.054	3.11	0.0020

Модель позволяет оценить стоимость квартиры с учетом рассмотренных выше факторов.

Вопросы для самопроверки

· Что такое эконометрика как наука?
· Что является предметом изучения эконометрики?
· Какой основной метод исследования в эконометрике?
· Какова цель эконометрического исследования?
· Каковы основные задачи эконометрики?
· Что такое эконометрическая модель?
· Каковы этапы эконометрического моделирования?
· Какие переменные применяются в эконометрическом анализе?
· Какие существуют типы эконометрических моделей?
· Что такое корреляционная зависимость?
· Что такое ковариация?
· Что такое коэффициент линейной корреляции?
· Что такое спецификация модели?
· Что такое идентификация эконометрической модели?

Дополнительная литература

· Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 2001. – 430 с. (глава 1, глава 2, п.2.1).
· Доугерти К. Введение в эконометрику. – М.: ИНФРА – М, 2009. – 465 с. (Обзор).
· Елисеева И.И. Эконометрика: Учебник / под. ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 344 с. (глава 1).

Интернет-ресурсы

· http://www.nsu.ru/ef/tsy/ecmr/index.htm
· http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html
· http://www.statsoft.ru/home/textbook/glossary/default.htm

Тема 2. Линейные и нелинейные модели парной регрессии

Оценка параметров парной линейной регрессии. Метод наименьших квадратов (МНК). Оценка значимости параметров регрессии и модели в целом. Точечный и интервальный прогноз по уравнению регрессии. Линеаризация нелинейной модели.

Задачи изучения:

· понять суть идентификации эконометрической модели и основную идею МНК,

· научиться оценивать параметры эконометрической модели по статистическим данным,

· научиться оценивать качество модели,

· научиться осуществлять прогнозирование по результатам моделирования,

· научиться преобразовывать нелинейные модели к линейным моделям.

Теоретический материал

Определение.

Парная регрессия представляет зависимость результативного признака только от одного факторного признака. Модель имеет вид:

$$y = \bar{y}(x) + \varepsilon$$

Подбор типа функции для построения выборочного уравнения регрессии в случае парной регрессии чаще всего осуществляется на основе графического представления выборочных данных.

Более точный анализ связан с получением нескольких моделей различных типов с последующим выбором наилучшей модели, более адекватно описывающей реальную связь признаков.

Типы функциональной зависимости:

- линейная $\bar{y} = ax + b$;

- квадратическая $\bar{y} = ax^2 + bx + c$;

- гиперболическая $\bar{y} = \frac{a}{x} + b$ и др.

a, b, c - параметры.

Критерии оптимальности модели.

Используются показатели, характеризующие суммарное отклонение выборочных значений результативного признака $\bar{y}_x$ от соответствующих значений $\hat{y}(x)$, рассчитанных по выборочному уравнению регрессии вида $\hat{y} = \hat{y}(x)$. К ним, в частности, относятся:

- *средняя ошибка аппроксимации* $\eta = \frac{1}{n}\sum_x \frac{|\bar{y}_x - \hat{y}(x)|}{\bar{y}_x} \cdot 100;$

- *остаточная дисперсия* $\sigma_\varepsilon = \frac{\sum_x (\hat{y}(x) - \bar{y}_x)^2}{n};$

- *сумма квадратов остатков* $Q = \sum_x (\hat{y}(x) - \bar{y}_x)^2.$

Определения и формулы.

Парная линейная регрессия характеризует линейную корреляционную зависимость Y от X.

Корреляционная зависимость:

$$\bar{y} = \alpha + \beta x$$

Оценка корреляционной зависимости (выборочное уравнение):

$$\hat{y} = \hat{\alpha} + \hat{\beta} x$$

Уравнение регрессии:

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

Оценка уравнения регрессии:

$$y = \hat{\alpha} + \hat{\beta} x + e$$

Теоретическое отклонение:

$$\varepsilon = y - (\alpha + \beta x)$$

Оценка теоретического отклонения (остаток или невязка регрессии):

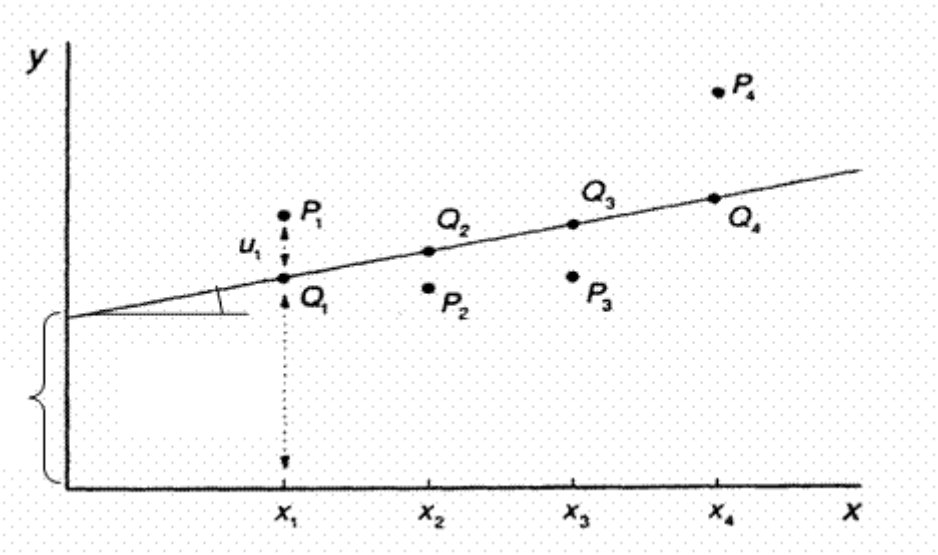
$$e = y - \left(\hat{\alpha} + \hat{\beta}\ x\right).$$

Выборочные значения параметров $\hat{\alpha}, \hat{\beta}$ являются точечными оценками параметров парной линейной регрессии соответственно α, β.

Величина β называется *коэффициентом линейной регрессии*. Она характеризует степень чувствительности результата от вариации фактора.

Оценки параметров находят методом наименьших квадратов по формулам:

a



,

$$\hat{\beta} == \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2} \qquad \hat{\alpha} = \frac{\bar{y} \cdot \overline{x^2} - \bar{x} \cdot \overline{xy}}{\sigma_x^2}$$

Вывод формул.

$$Q = \sum (\bar{y}_i - \hat{y}_i)^2 \to \min. \qquad \hat{y} = \hat{\alpha} + \hat{\beta} x$$

$$Q = \sum (\bar{y}_i - (\hat{\beta} x_i + \hat{\alpha}))^2 \to \min$$

$$\frac{\partial F}{\partial \hat{\alpha}} = -2\sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \qquad \frac{\partial F}{\partial \hat{\beta}} = -2\sum_{i=1}^{n} (y_i - \alpha_1 - \hat{\beta} x_i) x_i = 0, \text{ или}$$

$$\begin{cases} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \\ \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0 \end{cases} \quad \text{или} \quad \begin{cases} \sum_{i=1}^{n} e_i = 0 \\ \sum_{i=1}^{n} x_i e_i = 0 \end{cases}$$

$$\begin{cases} n\hat{\alpha} + (\sum_{i=1}^{n} x_i)\hat{\beta} = \sum_{i=1}^{n} y_i \\ (\sum_{i=1}^{n} x_i)\hat{\alpha} + (\sum_{i=1}^{n} x_i^2)\hat{\beta} = \sum_{i=1}^{n} y_i x_i \end{cases}$$

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{\beta} == \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2} \qquad \hat{\alpha} = \frac{\bar{y} \cdot \overline{x^2} - \bar{x} \cdot \overline{xy}}{\sigma_x^2}$$

Статистическое оценивание параметров регрессии.

Для проверки гипотез о значимости $\hat{a}, \hat{b}$ используются критерии Стьюдента, выборочные значения которых вычисляются по формулам:

$$t_a = \hat{a}\frac{\sqrt{n-2}s_x}{s_\varepsilon},$$

$$t_B = \hat{b}\frac{\sqrt{n-2}}{s_\varepsilon},$$

где s_x - оценка среднего квадратического отклонения σ_x выборочных значений факторного признака от выборочной средней, s_ε - оценка среднего квадратического отклонения σ_ε выборочных значений результативного признака от соответствующих им теоретических значений, вычисленных с учетом уравнения регрессии:

$$\sigma_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}}, \quad \sigma_\varepsilon = \sqrt{\frac{\sum_i (\hat{y}(x_i) - \bar{y}_{x_i})^2}{n}},$$

$$s_x = \sqrt{\frac{\sum_i (x_i - \bar{x}^2)^2}{n-1}}, \quad s_\varepsilon = \sqrt{\frac{\sum_i (\hat{y}(x_i) - \bar{y}_{x_i})^2}{n-2}}.$$

Далее делаются выводы: если выборочные значения параметров по абсолютной величине больше критического значения критерия Стьюдента при заданном уровне значимости, то соответствующие параметры признаются значимыми, а модель – пригодной для практического использования. В противном случае производятся дополнительные исследования, в частности, связанные с увеличением объема выборочных данных.

Определение интервальных оценок параметров модели производится стандартным образом по формулам:

$$a \in (\hat{a} - t_\gamma \sigma_a, \hat{a} + t_\gamma s_a), \quad b \in (\hat{b} - t_\gamma \sigma_b, \hat{a} + t_\gamma s_b),$$

где s_a, s_b - точечные оценки средних квадратических отклонений значений параметров по выборочным данным:

$$s_a = \frac{s_\varepsilon}{s_x\sqrt{n-2}}, \quad s_b = \frac{s_\varepsilon}{\sqrt{n-2}}.$$

Оценка качества уравнения в целом.

Оценка значимости уравнения регрессии в целом производится на основе F - критерия Фишера, которому предшествует дисперсионный анализ. В математической статистике дисперсионный анализ рассматривается как самостоятельный инструмент

статистического анализа. В эконометрике он применяется как вспомогательное средство для изучения качества регрессионной модели.

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

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

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

Схема дисперсионного анализа имеет вид, представленный в таблице (n – число наблюдений, m – число параметров при переменной x).

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

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F - критерия Фишера:

$$F = \frac{S^2_{факт}}{S^2_{ост}}.$$

Фактическое значение F -критерия Фишера сравнивается с табличным значением $F_{табл}(\alpha; k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F -критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m = 1$, поэтому

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

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

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2).$$

Прогнозирование.

Построенная регрессионная модель применяется для прогнозирования результата при заданном значении
фактора x_0. Точечная оценка индивидуального прогнозного значения определяется по формуле:

$$y(x_0) \approx \hat{y}(x_0) = \hat{a}x_0 + \hat{b}.$$

Доверительный интервал для среднего значения $\bar{y}_{x_0}$ находят по формуле:

$$\bar{y}_{x_0} \in (\hat{y}(x_0) - t_\gamma s_{\hat{y}}, \hat{y}(x_0) + t_\gamma s_{\hat{y}}),$$

где величина $s_{\hat{y}}$ является точечной оценкой среднего квадратического отклонения прогнозного значения результата:

$$s_{\hat{y}} = s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}}.$$

Доверительный интервал для оценки индивидуального значения результата $y(x_0)$ определяется с учетом вариации значения результативного признака при фиксированном значении фактора:

$$y(x_0) \in (\hat{y}(x_0) - t_\gamma s_y, \hat{y}(x_0) + t_\gamma s_y),$$

где s_y - оценка общей вариации результата, обусловленной действием случайных факторов ε, а также ошибками выборочного исследования уравнения регрессии:

$$s_y = s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_0 - x)^2}{(n-1)s_x}}.$$

Линеаризация моделей.

Для приведения нелинейных моделей к линейному виду используют процедуры замены переменных и логарифмирования. Далее приведены примеры линеаризации наиболее распространенных функций.

· Гиперболическая функция:

$$\bar{y} = \beta_0 + \beta_1 \frac{1}{x}$$

Сделаем замену переменных:

$$X = \frac{1}{x},$$

уравнение примет вид:

$$\bar{y} = \beta_0 + \beta_1 X.$$

· Полулогарифмическая функция:

$$\bar{y} = \beta_0 + \beta_1 \ln x$$

Сделаем замену переменных:

$$X = \ln x,$$

уравнение примет вид:

$$\bar{y} = \beta_0 + \beta_1 X.$$

· Обратная функция:

$$\bar{y} = \frac{1}{\beta_0 + \beta_1 x}$$

Сделаем замену переменных:

$$\bar{Y} = \frac{1}{\bar{y}},$$

уравнение примет вид:

$$\bar{Y} = \beta_0 + \beta_1 x.$$

· Показательная функция:

$$\bar{y} = \beta_0 \ell^{\beta_1 x}$$

Прологарифмируем уравнение:

$$\ln \bar{y} = \ln \beta_0 + \beta_1 x,$$

сделаем замену переменных:

$$\bar{Y} = \ln \bar{y}, \quad B_0 = \ln \beta_0,$$

уравнение примет вид:

$$\bar{Y} = B_0 + \beta_1 x$$

· Степенная функция:

$$\bar{y} = \beta_0 x^{\beta_1}$$

Прологарифмируем уравнение:

$$\ln \bar{y} = \ln \beta_0 + \beta_1 \ln x,$$

сделаем замену переменных:

$$\bar{Y} = \ln \bar{y}, \quad X = \ln x, \quad B_0 = \ln \beta_0,$$

уравнение примет вид:

$$\bar{Y} = B_0 + \beta_1 X.$$

Пример. Исследование зависимости розничного товарооборота магазинов от среднесписочного числа работников.

В таблице приведены данные по 8 магазинам. *x* – численность работающих, (чел.), *y* – величина розничного товарооборота (млн. руб.).

№ п/п	***x***	***y***	$\hat{y}$
1	73	0,5	0,43
2	85	0,7	0,661
3	102	0,9	0,998
4	115	1,1	1,239
5	122	1,4	1,373
6	126	1,4	1,45
7	134	1,7	1,604
8	147	1,9	1,854

Средние значения показателей:

$$\bar{x} = 113 \quad \bar{y} = 1,2$$

Вспомогательные значения для определения параметров регрессионной модели:

$$\overline{x^2} = 13313,5 \quad \overline{y^2} = 1,65 \quad \overline{xy} = 146,075$$

Показатели вариации показателей:

$$\sigma_x^2 = 13313,5 - 113^2 = 544,5 \quad s_x^2 = \frac{8}{7}544,5 = 622,286 \quad s_x = \sqrt{622,286} = 24,95$$

$$\sigma_y^2 = 1,65 - 1,2^2 = 0,21 \quad s_y^2 = \frac{8}{7}0,21 = 0,24 \quad s_y = \sqrt{0,24} = 0,490$$

Оценки параметров парной линейной регрессии:

$$\hat{a} = \frac{n\sum x\bar{y}_x - \sum x\sum \bar{y}_x}{n\sum x^2 - (\sum x)^2} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{146,075 - 113\cdot 1,2}{13313,5 - 113^2} = 0,019$$

$$\hat{b} = \frac{\sum \bar{y}_x \sum x^2 - \sum x\sum x\bar{y}_x}{n\sum x^2 - (\sum x)^2} = \frac{\overline{x^2}\cdot\bar{y} - \bar{x}\cdot\overline{xy}}{\overline{x^2} - (\bar{x})^2} = \frac{13313,5\cdot 1,2 - 113\cdot 146,075}{13313,5 - 113^2} = -0,974$$

Выборочное уравнение регрессии:

$$\hat{y} = -0,974 + 0,019x$$

Интерпретация модели:

При увеличении численности занятых на одного работника величина товарооборота возрастет на 19 тысяч рублей. Свободный член в модели не имеет экономического смысла (он равен здесь величине товарооборота при нулевой численности работников).

Оценки вариации параметров уравнения регрессии:

$$s_\varepsilon = \sqrt{\frac{\sum_i (\hat{y}(x_i) - \bar{y}_{x_i})^2}{n-2}} = \sqrt{\frac{(0,5-0,43)^2 + \ldots + (1,9-1,854)^2}{8-2}} = 0,089$$

$$s_a = \frac{s_\varepsilon}{s_x\sqrt{n-2}} = \frac{0,089}{24,95\cdot\sqrt{8-2}} = 0,0015$$

$$s_b = \frac{s_\varepsilon}{\sqrt{n-2}} = \frac{0,089}{\sqrt{8-2}} = 0,036$$

Расчетные значения статистики Стьюдента:

$$t_a = \hat{a}\frac{\sqrt{n-1}s_x}{s_\varepsilon} = \frac{\hat{a}}{s_a} = \frac{0,019}{0,0015} = 12,67$$

$$t_B = \hat{b}\frac{\sqrt{n-2}}{s_\varepsilon} = \frac{\hat{b}}{s_b} = \frac{-0,974}{0,036} = -27,06$$

Оба коэффициента значимы при $\alpha = 5\%$. Это означает, что ошибаясь в 5 случаях из 100, можно утверждать, что связь между x и y существенна.

Коэффициент детерминации:

$$B = R^2 = 1 - \frac{S_\varepsilon^2}{S_y^2} = 1 - \frac{0,089^2}{0,24} = 0,97.$$

Это означает, что вариация товарооборота на 97% процентов обусловлена численностью работников и только на 3% - остальными факторами.

Интервальные оценки для коэффициентов при $\gamma = 0,95$:

$$a \in (0,019 \pm 2,45 \cdot 0,0015) \quad a \in (0,015; 0,022)$$

$$b \in (-0,974 \pm 2,45 \cdot 0,036) \quad b \in (-1,062; -0,886)$$

Это означает, что истинные значения коэффициентов в модели с вероятностью 95% лежат в указанных пределах.

Используем модель для прогнозирования. Найдем оценку прогнозного значения товарооборота для численности работников 140 человек.

Точечная оценка прогноза:

$$\hat{y}(140) = -0,974 + 0,019 \cdot 140 = 1,69$$

Стандартная ошибка среднего значения прогноза:

$$S_{\hat{y}} = S_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 0,089\sqrt{\frac{1}{8} + \frac{(140-113)^2}{4357,5}} = 0,048$$

Интервальная оценка среднего значения прогноза:

$$\bar{y}(140) \in (1,69 \pm 2,45 \cdot 0,048) \quad \text{или} \quad \bar{y}(140) \in (1,572; 1,81)$$

Стандартная ошибка индивидуального значения прогноза:

$$S_{y(140)}^2 = 0,048^2 + 0,089^2 = 0,010$$

$$s_{y(140)} = \sqrt{0,010} = 0,101$$

Интервальная оценка индивидуального значения прогноза:

$$y(140) \in (1,69 \pm 2,45 \cdot 0,101) \text{ или } y(140) \in (1,443;1,897).$$

Вопросы для самопроверки

· Что представляет собой парная регрессионная линейная модель?
· В чем суть МНК?
· Что значит оценить значимость параметров уравнения регрессии?
· Каков алгоритм оценки значимости параметров парной линейной регрессии?
· Что значит оценить качество уравнения регрессии в целом?
· Каков алгоритм оценки качества парной регрессионной модели?
· Что такое коэффициент детерминации?
· Как использовать регрессионную модуль для прогнозирования?
· Каков алгоритм прогнозирования с использованием парной линейной регрессионной модели?
· Какие существуют парные нелинейные регрессионные модели?
· Какие существуют способы приведения нелинейных регрессионных моделей к линейным моделям?

Дополнительная литература

· Айвазян С.А. Иванова С.С. Эконометрика. Краткий курс: учеб. пособие. – М.: Маркет ДС, 2007. – 104 с. (глава 1, п. 1.1).
· Доугерти К. Введение в эконометрику. – М.: ИНФРА – М, 2009. – 465 с. (Главы 1, 2,4).
· Елисеева И.И. Эконометрика: Учебник / под. ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 344 с. (глава 2).

Интернет-ресурсы

· http://www.nsu.ru/ef/tsy/ecmr/index.htm
· http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html
· http://www.statsoft.ru/home/textbook/glossary/default.htm

Тема 3. Модели множественной регрессии

Подбор факторов множественной регрессии. Оценка параметров и их значимости уравнения множественной линейной регрессии. Точечный и интервальный прогноз по уравнению регрессии. Фиктивные переменные.

Задачи изучения:

· научиться отбирать факторы для модели множественной регрессии,
· научиться оценивать параметры модели множественной регрессии,
· научиться оценивать качество модели множественной регрессии,
· научиться использовать модель множественной регрессии для прогноза,
· научиться строить модель с фиктивными переменными.

Теоретический материал

Выбор факторов.

В большинстве случаев существенное влияние на результат оказывают несколько факторов. Модель множественной регрессии, характеризующая зависимость между тремя и более признаками имеет вид:

$$y = \bar{y}(x_1, x_2, \ldots, x_k) + \varepsilon.$$

Функция $\bar{y}(x_1, x_2, \ldots, x_k)$ корреляционную зависимость признака Y от факторов $X_1, X_2, \ldots, X_k$.

Построение моделей множественной регрессии включает следующие взаимосвязанные задачи:

- отбор факторных признаков;
- выбор формы связи;
- статистическое оценивание параметров уравнения регрессии;
- проверка адекватности модели.

Для решения проблемы отбора факторных признаков используют следующие методы:

- *метод экспертных оценок*, основанный на интуитивно-логических предпосылках и содержательно-качественном анализе информации с привлечением специальных экспертов;
- *метод корреляции,* базирующийся на анализе выборочных значений показателей связи различных факторов;
- *метод шаговой регрессии*, который заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости.

Критериями отбора факторов X_i, X_j методом корреляции являются следующие соотношения:

$$r_{yi} > r_{ij}, \quad r_{yj} > r_{ij}, \quad r_{ij} < 0.8,$$

где r_{yi}, r_{yj} - коэффициенты корреляции между результатом и каждым из факторов, r_{ij} - коэффициент корреляции между факторами.

Невыполнение последнего неравенства свидетельствует о наличии явления мультиколлинеарности - тесной связи между факторными признаками, которое приводит к искажению величин параметров модели. Устранение явления мультиколлинеарности реализуют путем устранения одного из факторов, либо их объединения в один общий фактор.

Шаговая регрессия является наиболее приемлемым способом отбора факторных признаков. При проверке значимости очередного введенного фактора определяется, насколько уменьшается сумма квадратов остатков и увеличивается величина коэффициента множественной корреляции. Фактор считается несущественным, если:

- его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не изменяя суммы квадратов остатков;
- коэффициенты регрессии меняют не только величину, но и знаки, а множественный коэффициент корреляции не возрастает;
- на основе результатов статистического оценивания поверки значимости.

Фактор считается существенным, если увеличивается значение множественного коэффициента корреляции при неизменном коэффициенте регрессии.

Выбор формы связи осуществляется перебором моделей с учетом показателей меры отклонений эмпирических и теоретических данных, как и в случае парной регрессии.

Линейные модели множественной регрессии.

Наиболее распространены *линейные* модели множественной регрессии. Они имеют вид:

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_kx_k + \varepsilon.$$

$\bar{y} = a_0 + a_1x_1 + a_2x_2 + \ldots + a_kx_k$ - детерминированная составляющая.

$$y = \bar{y} + \varepsilon$$

$\hat{y} = \hat{a}_0 + \hat{a}_1x_1 + \hat{a}_2x_2 + \ldots + \hat{a}_kx_k$ - выборочное уравнение регрессии.

$y = \hat{y} + e$ - регрессионная модель, найденная по выборочным данным.

Оценка параметров выборочного уравнения регрессии производится на основе метода наименьших квадратов, применяемого в матричном виде.

Вывод формулы для оценок коэффициентов регрессии.

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

В матричном виде:

$$Q = e^T e = (y - \hat{y})^T (y - \hat{y}) = (y^T - (X\hat{a})^T)(y - (X\hat{a})) = y^T y - y^T (X\hat{a}) - (X\hat{a})^T y + (X\hat{a})^T (X\hat{a}) =$$
$$= y^T y - 2y^T (X\hat{a}) + \hat{a}^T X^T X\hat{a}$$

$$\frac{dQ}{d\hat{a}} = -2y^T X + 2\hat{a}^T X^T X = 0$$

Нормальное уравнение:

$$y^T X - \hat{a}^T X^T X = 0$$

$$\hat{a}^T = (X^T X)^{-1} y^T X$$

Окончательно:

$$\hat{a} = (X^T X)^{-1} X^T y,$$

где величины $\hat{a}$, X, y представляют матричную форму записи значений параметров и признаков, определенных по n выборочным данным:

$$\hat{a} = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \dots \\ \hat{a}_k \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}.$$

Значения y_i представляют собой средние значения признака Y по результатам i-го наблюдения при фиксированных значениях всех учитываемых факторов:

$$y_i = \bar{y}_{x_{i1}, x_{i2}, \dots, x_{ik}}.$$

Коэффициент детерминации и скорректированный коэффициент детерминации

Коэффициентом детерминации

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}.$$

R^2 возрастает при добавлении еще одного регрессора, поэтому для выбора между несколькими регрессионными уравнениями не следует полагаться только на R^2.

Попыткой устранить эффект, связанный с ростом R^2 при увеличении числа регрессоров, является коррекция R^2 на число регрессоров - наложение "штрафа" за увеличение числа независимых переменных.

Скорректированный R^2

$$R^2_{adj} = 1 - \frac{ESS/(N-k)}{TSS/(N-1)}.$$

Здесь в числителе - несмещенная оценка дисперсии ошибок, в знаменателе - несмещенная оценка дисперсии *Y*.

Свойства скорректированного R^2:

1. $R^2_{adj} = 1 - (1 - R^2)\frac{N-1}{N-k}$;

2. $R^2 > R^2_{adj}$;

3. $R^2_{adj} \le 1$

Использование R^2_{adj} для сравнении регрессий при изменении числа регрессоров более корректно.

Оценка качества модели.

Проблема практической пригодности моделей множественной регрессии связана с решением двух взаимосвязанных задач:

- статистическое оценивание параметров уравнения регрессии;
- проверка гипотезы о несоответствии заложенных в уравнение регрессии и реально существующих связей между признаками.

В соответствии с решением этих задач возможны следующие варианты выводов о приемлемости модели:

- если все параметры значимы и сформулированная гипотеза отвергается, то модель считается пригодной для принятия решений;
- если часть параметров незначима и гипотеза отвергается, то модель неприменима при решении задачи прогнозирования, однако может быть использована в экономическом анализе путем интерпретации отдельных ее параметров;
- если все параметры незначимы, то модель считается непригодной для практического использования.

Оценка значимости параметров регрессии производится с использованием критерия Стьюдента в виде:

$$t_B = \frac{\hat{a}_i}{s_{a_i}}, \quad i = 0, 1, 2, \ldots, k.$$

Величина s_{a_i} является оценкой среднего квадратического для $\hat{a}_i$:

$$s_{a_i} = S^2\sqrt{b_{ii}},$$

где b_{ii} - диагональные элементы матрицы $(X^TX)^{-1}$, S^2 - оценка среднего квадратического остатков:

$$S^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}.$$

Доверительные интервалы параметров регрессии находят по формулам:

$$a_i \in (\hat{a}_i - t_\gamma s_{a_i}, \hat{a}_i + t_\gamma s_{a_i}).$$

Анализ адекватности модели осуществляется как проверка гипотезы о несоответствии заложенных в уравнение и реально существующих связей. Используется статистический критерий Фишера:

$$f_B = \frac{\sum_{i=1}^{n}\hat{y}_i^2}{(k+1)S^2}.$$

С целью расширения возможностей экономического анализа используются *частные коэффициенты эластичности*, определяемые по формулам:

$$E_i = \hat{a}_i \frac{\bar{x}_i}{\bar{y}} \cdot 100,$$

где $\bar{x}_i, \bar{y}$ - средние выборочные значения признаков X_i, Y. Коэффициент эластичности показывает, на сколько процентов в среднем изменится значение результативного признака при изменении i -го фактора на один процент.

Прогнозирование.

Доверительный интервал прогноза находят по формуле:

$$y \in (\hat{y} - t_\gamma s_{\hat{y}}, \hat{y} + t_\gamma s_{\hat{y}}),$$

$$s_{\hat{y}} = s\sqrt{X_0^T(X^TX)^{-1}X_0},$$

где $X_0 = (1, x_{01}, x_{02}, \ldots, x_{0k})$ - вектор заданных значений факторов.

Фиктивные переменные.

Может оказаться необходимым включить в модель фактор, имеющий два или более качественных уровней. Это могут быть разного рода атрибутивные признаки, такие, например, как профессия, пол, образование, климатические условия, принадлежность к определенному региону. Чтобы ввести такие переменные в регрессионную модель, им должны быть присвоены те или иные *цифровые метки,*

т.е. качественные переменные преобразованы в количественные. Такого вида сконструированные переменные в эконометрике принято называть *фиктивными переменными*.

Вопросы для самопроверки

· Что представляет собой модель множественной регрессии?
· Каковы проблемы подбора факторов в модели множественной регрессии?
· Каков алгоритм подбора факторов в множественной регрессии?
· Какие существуют методы оценки параметров множественной регрессии?
· Каков алгоритм применения МНК для оценки параметров множественной линейной регрессии?
· Каков алгоритм оценки значимости параметров множественной линейной регрессии?
· Каков алгоритм оценки качества уравнения множественной линейной регрессии в целом?
· Каков алгоритм прогнозирования с использованием модели множественной линейной регрессии?
· Что означают фиктивные переменные?
· Каков алгоритм применения фиктивных переменных в моделях множественной линейной регрессии?

Дополнительная литература

· Айвазян С.А. Иванова С.С. Эконометрика. Краткий курс: учеб. пособие. – М.: Маркет ДС, 2007. – 104 с. (глава 1, п. 1.2, глава 2, п.2.1).
· Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 2001. – 430 с. (глава 1, глава 2, п. 2.2, п.2.3).
· Доугерти К. Введение в эконометрику. – М.: ИНФРА – М, 2009. – 465 с. (Главы 3, 5).
· Елисеева И.И. Эконометрика: Учебник / под. ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 344 с. (глава 3, п.3.1- п.3.9).

Интернет-ресурсы

· http://www.nsu.ru/ef/tsy/ecmr/index.htm
· http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html
· http://www.statsoft.ru/home/textbook/glossary/default.htm

Тема 4. Предпосылки метода наименьших квадратов

Гетероскедастичность остатков. Автокорреляция остатков. Теорема Гаусса-Маркова. Обобщенный метод наименьших квадратов (ОМНК). Теорема Айткена.

Стохастические регрессоры. Метод инструментальных переменных (МИП). Мультиколлинеарность факторов.

Задачи изучения темы:

· научиться выявлять нарушение предпосылок МНК,

· научиться устранять проявление нарушения предпосылок МНК.

Теоретический материал

Для того чтобы оценки, полученные по МНК, давали «наилучшие» результаты, мы потребуем от остаточного члена или ошибки e и от X выполнения следующих условий.

1. $Y = \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$ - спецификация модели.

Отражает представление о механизме зависимости Y и X и выбор объясняющей переменной X.

2. $X_1,\ldots,X_k$ – детерминированные вектора, линейно независимые в R^n, т. е. матрица X имеет ранг k.

Линейная независимость нужна для совместности системы нормальных уравнений. В случае зависимости определитель системы мал и вносит большую погрешность.

Из детерминированности следует условие, более сильное, что объясняющие переменные не коррелируют со случайной переменной.

Во-первых, ŷ коррелирует с ?; во-вторых, возможна связь X и Y, т.е. взаимосвязь. Это невозможно, так как регрессии X на Y и Y на X совпадают при функциональной зависимости.

3. $M\varepsilon_i = 0$.

Это обуславливает предположение, что при МНК предполагается, что y зависит только от x:

$M[y\varepsilon] = M[(\bar{y} + \varepsilon)\varepsilon] = M[\bar{y}\varepsilon + \varepsilon^2] = M[\bar{y}\varepsilon] + M[\varepsilon^2] = M[\bar{y}]M[\varepsilon] + M[\varepsilon^2] = 0$ тогда и только тогда, когда $M\varepsilon = 0$.

4. $M\varepsilon_i^2 = D\varepsilon_i = \sigma_\varepsilon^2$, дисперсия ошибки не зависит от номера наблюдения.

Условие независимости ошибок от номера наблюдения называют *гомоскедастичностью*. Случай, когда условие гомоскедастичности нарушается, называется *гетероскедастичностью*. Это означает, что в каждом наблюдении неучтенные факторы оказывают одинаковое влияние.

5. $M(\varepsilon_i\varepsilon_j) = 0$ при $i \neq k$, т. е. некоррелированность ошибок разных наблюдений.

Предполагает отсутствие систематической связи между значениями случайного члена в любых двух наблюдениях. Почти всегда нарушается, если данные представляют собой временные ряды. Если это условие не выполняется, говорят об *автокорреляции остатков.*

Отсутствие автокорреляции означает, что все существенные переменные уже учтены в x. Если бы это было бы не так, то y зависел от ?. Если $\varepsilon_j = f(\varepsilon_i)$, то ε_j -

существенный фактор.

6. $\varepsilon_i \in N(0, \sigma_\varepsilon)$.

Это следует из того, что e_i включает в себя много факторов, которые можно считать независимыми и нужно для получения интервальных оценок.

Теорема Гаусса-Маркова.

В условиях 1-5 МНК-оценки представляют собой наилучшие линейные несмещенные оценки. При выполнении условия оценки и регрессия распределены нормально: $\hat{a}_i \in N(a, \sigma_{\hat{a}_i})$ $\hat{y} \in N(\bar{y}, \sigma_{\hat{y}})$

Автокорреляция остатков.

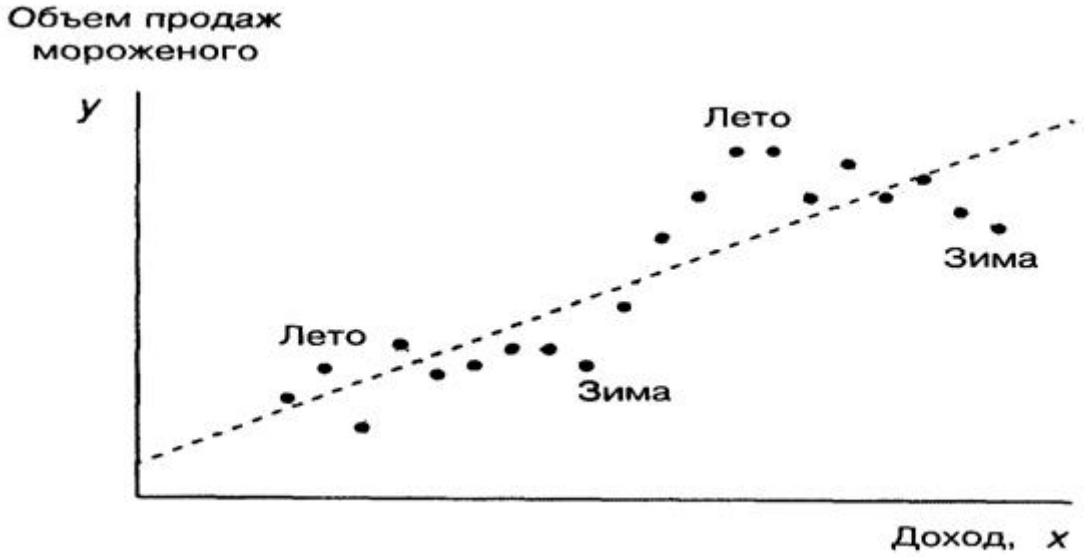
Причины автокорреляции

· Не учтена важная объясняющая переменная.

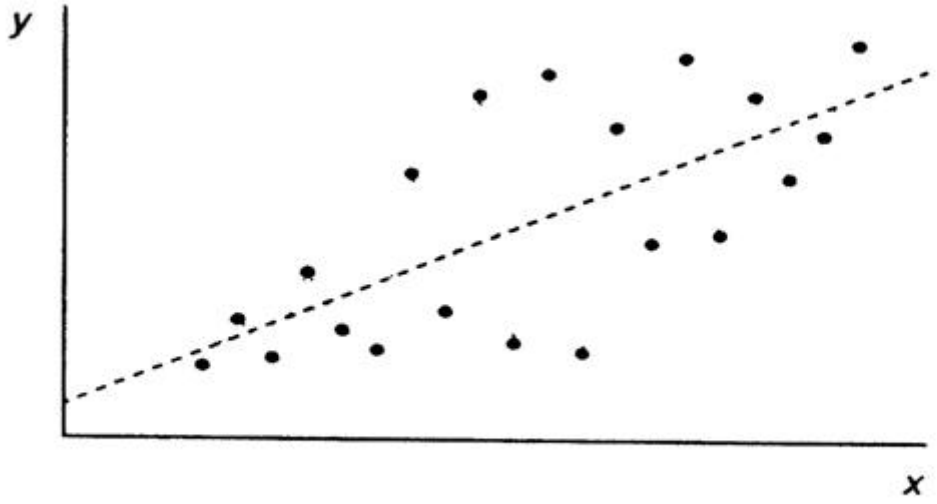
· Неадекватная функция регрессии.

· Числовой материал содержит большие ошибки наблюдений.

Обнаружение автокорреляции остатков производится путем графического анализа остатков и использования критерия Дарбина-Уотсона.



Положительная автокорреляция



Отрицательная автокорреляция

Критерий Дарбина-Уотсона.

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

$\rho_{e_t e_{t-1}} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$ - коэффициент линейной корреляции

$$d \approx 2(1 - \rho_{e_t e_{t-1}})$$

$d = 2 \quad \rho_{e_t e_{t-1}} = 0$ - отсутствует корреляция

$d = 0 \quad \rho_{e_t e_{t-1}} = 1$ - положительная корреляция

$d = 4 \quad \rho_{e_t e_{t-1}} = -1$ - отрицательная корреляция

Величина d зависит от значений факторов, поэтому однозначно критическое значение найти нельзя. Находят верхнюю и нижнюю границы:

$$d_L(\alpha, n, k), \quad d_U(\alpha, n, k)$$

Правило:

$d_U \le d \le 4 - d_U$ - принимается гипотеза H_0 об отсутствии автокорреляции остатков.

$d \approx 2 \quad \rho_{..} \approx 0$

$\rho_{e_t e_{t-1}}$

$0 \le d \le d_L$ - принимается гипотеза H_1 о наличии положительной автокорреляции остатков.

$d \approx 0 \quad \rho_{e_t e_{t-1}} \approx 1$

$d_L \le d \le d_U \quad \vee \quad 4 - d_U \le d \le 4 - d_L$ - при выбранном уровне значимости нельзя прийти к определенному выводу.

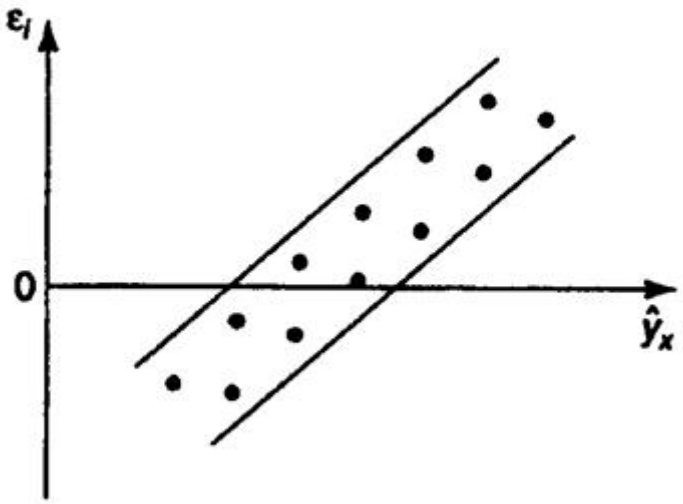
$4 - d_L \le d \le 4$ - принимается гипотеза H_1 о наличии отрицательной автокорреляции остатков.

$d \approx 4 \quad \rho_{e_t e_{t-1}} \approx -1$

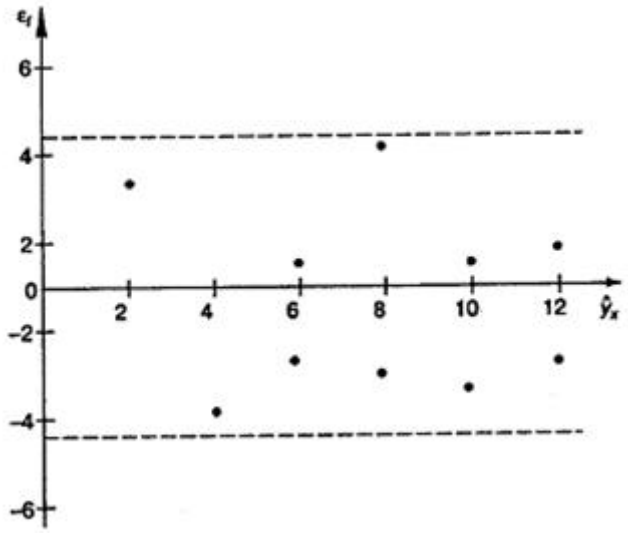
Гетероскедастичность остатков.

Гетероскедастичность остатков означает их неоднородность и количественно выражается в зависимости дисперсии остатков от факторов.

Способ обнаружения гетероскедастичности основан на графическом анализе остатков и использовании техники проверки гипотезы о неизменности дисперсии. При этом используются критерии Гольдфельда-Квандта, Уайта, Глейзера идр.



Остатки гетероскедастичны, поскольку они зависят от линейной комбинации факторов.



Остатки гомоскедастичны, поскольку они не зависят от линейной комбинации факторов.

Тест Гольдфельда-Квандта.

· всю выборку делят на 3 части,

· оценивают регрессии для крайних выборок,

· сравнивают оценки дисперсии остатков для полученных регрессий:

· если эти оценки отличаются незначимо, то делают вывод об отсутствии гетероскедастчности остатков в исходной модели,

· в противном случае гипотеза об отсутствии гетероскедастичности остатков отклоняется.

В случаях обнаружения гетероскедастичности и автокорреляции остатков используется обобщенный метод наименьших квадратов (ОМНК), обоснование которому дает теорема Айткена.

Теорема Айткена.

Рассматривается обобщенная линейная модель

$$Y = X\beta + \varepsilon,$$

в которой нет ограничений на отсутствие автокорреляции и гетероскедастичности остатков.

В классе несмещенных линейных (по Y) оценок вектора b для обобщенной регрессионной линейной модели оценка $\hat{\beta}_* = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}Y$ имеет наименьшую матрицу ковариаций (Ω - матрица ковариаций остатков).

В практических расчетах используется так называемый доступный ОМНК, в котором в качестве матрицы ковариаций выступает ее оценка, полученная после применения обычного МНК.

Анализ мультиколлинеарности.

Мультиколлинеарность означает зависимость факторов. Ее следует избежать на этапе отбора факторов путем анализа матрицы парных коэффициентов корреляции:

$$\begin{pmatrix} 1 & r_{y1} & r_{y2} & \dots & r_{yk} \\ r_{1y} & 1 & r_{12} & \dots & r_{1k} \\ r_{2y} & r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{ky} & r_{k1} & r_{k2} & \dots & 1 \end{pmatrix},$$

Анализ таблицы ведется с использованием следующих критериев:

1. $|r_{yj}| > |r_{ij}|, \quad |r_{yi}| > |r_{ij}|, \quad |r_{ij}| < 0,8, \quad i,j = 1,2,\dots k$

2. $r_{yj} > r_{yi}$

Стохастические регрессоры.

Если факторы являются случайными величинами и коррелируют со случайными возмущениями, то оценки МНК будут смещенными и, возможно, несостоятельными. В этом случае для идентификации модели применяют метод инструментальных переменных.

Метод инструментальных переменных.

Постановка задачи.

$$Y = X\beta + \varepsilon$$

Требуется подобрать такие инструментальные переменные (ИП) *Z*, чтобы они хорошо коррелировали с X и не коррелировали с e.

Оценкой b с помощью ИП называется оценка вида $\hat{\beta}_{ип} = (Z^T X)^{-1} Z^T Y,$ где $Z = \|z_{ij}\|_{n \times m}.$

При этом предполагается, что

$\frac{1}{n} Z^T X \xrightarrow[n\to\infty]{p} E[Z^T X], \quad |E[Z^T X]| \neq 0$ - характеризует хорошую коррелируемость X и Z.

$cov[Z_l, \varepsilon] = 0, \quad l = 1,2,\dots,m.$

Вопросы для самопроверки

· В чем состоит суть проблемы предпосылок применения МНК?

· Каковы предпосылки применения МНК для парной линейной регрессии?

· Каковы предпосыки применения МНК для множественной линейной регрессии?

· Что такое гетероскедастичность остатков?
· Что такое автокорреляция остатков?
· Что такое мультиколлинеарность факторов?
· В чем смысл теоремы Гаусса-Маркова?
· Что такое ОМНК?
· В чем смысл теоремы Айткена?
· Что такое стохастические регрессоры?
· Что такое инструментальные переменные?
· Каков алгоритм применения МИП?

Дополнительная литература

· Айвазян С.А. Иванова С.С. Эконометрика. Краткий курс: учеб. пособие. – М.: Маркет ДС, 2007. – 104 с. (глава 2, п. 2.2, глава 3, 4, 5, 7).
· Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 2001. – 430 с. (глава 2, п. 2.4 -2.13).
· Доугерти К. Введение в эконометрику. – М.: ИНФРА – М, 2009. – 465 с. (Главы 6-8).
· Елисеева И.И. Эконометрика: Учебник / под. ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 344 с. (глава 3, п.3.10 – п.3.11).

Интернет-ресурсы

· http://www.nsu.ru/ef/tsy/ecmr/index.htm
· http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html
· http://www.statsoft.ru/home/textbook/glossary/default.htm

Тема 5. Системы эконометрических уравнений

Структурная и приведенная формы модели. Виды систем эконометрических уравнений. Идентификация. Косвенный и двухшаговый метод наименьших квадратов.

Задачи изучения:
· научиться строить модели в виде систем одновременных уравнений,
· научиться преобразовывать модели из структурной формы в приведенную и наоборот,
· научиться определять идентифицируемость уравнений модели.

Теоретический материал

Внешне не связанные уравнения.

Связаны только благодаря корреляции остатков.

Постановка задачи.

$$\begin{cases} Y_1 = X_1\beta_1 + \varepsilon_1 \\ Y_2 = X_2\beta_2 + \varepsilon_1 \\ \dots \\ Y_m = X_m\beta_m + \varepsilon_m \end{cases}$$

$$Y_i = (Y_{1i}, Y_{2i}, \dots, Y_{ni})^T, \quad X = \|x_{ij}\|_{n\times k_i}, \quad \beta_i = (\beta_{1i}, \beta_{2i}, \dots, \beta_{k_i i})^T, \quad \varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i}, \dots, \varepsilon_{ni})^T$$

$$E[\varepsilon_i] = 0,$$

$$\Sigma[\varepsilon] = \sigma_{ij} I_n, \quad (i, j = 1, 2, \dots, m), \quad \sigma_{ij} = \operatorname{cov}[\varepsilon_{it}, \varepsilon_{js}] = \begin{cases} \neq 0, & t = s \\ 0, & t \neq s \end{cases}, \quad (t, s = 1, 2, \dots, n),$$

Последнее означает, что ошибки коррелируют только в одном наблюдении.
Каждое уравнение удовлетворяет предпосылкам МНК.

Оценка параметров.
Можно использовать МНК отдельно для каждого уравнения.
Можно улучшить оценки, объединив уравнения и используя связь между уравнениями.

$$Y = X\beta + \varepsilon$$

$$Y = (Y_1, Y_2, \dots, Y_m)^T, \quad X = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_m \end{pmatrix}, \quad \beta = (\beta_1, \beta_2, \dots, \beta_m)^T, \quad \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)^T,$$

$$\Omega = \Sigma[\varepsilon] \otimes I_n, \qquad A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix}.$$

Применим ОМНК

$$\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y$$

Оценка Ω

$$\text{МНК} \Rightarrow \quad \sigma_{ij} = \frac{e_i^T e_j}{n} \quad \Rightarrow \quad \hat{\Sigma} \quad \Rightarrow \quad \hat{\Omega} \otimes I_n$$

Связь с МНК.

Оценки ОМНК совпадают с оценками МНК в следующих случаях

1. $\sigma_{ij} = 0, \quad i \neq j$ - отсутствует корреляция остатков – уравнения не связаны.

2. $X_1 = X_2 = \ldots = X_m$ - в каждом уравнении одинаковые наборы экзогенных переменных.

Системы одновременных регрессионных уравнений.

Постановка задачи.

$BY + \Gamma X = \varepsilon$ - структурная форма модели

$B = \|\beta_{ij}\|_{m\times m}$, $\Gamma = \|\gamma_{ij}\|_{m\times k}$, $Y = (Y_1, Y_2, \ldots, Y_m)^T$, $X = (X_1, X_2, \ldots, X_k)^T$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)^T$.

1. $E[\varepsilon] = 0,$

2. $\Sigma[\varepsilon] = \sigma_{ij} I_n, \quad (i, j = 1, 2, \ldots, m), \quad \sigma_{ij} = \mathrm{cov}[\varepsilon_{it}, \varepsilon_{js}] = \begin{cases} \neq 0, & t = s \\ 0, & t \neq s \end{cases}$, $(t, s = 1, 2, \ldots, n).$

3. $\mathrm{cov}[X, \varepsilon] = 0$

4. $|B| \neq 0$

5. Условие нормировки: один из коэффициентов при Y в каждом уравнении равен единице.

$B^{-1} \mid \quad BY + \Gamma X = \varepsilon$

$Y + B^{-1}\Gamma X = B^{-1}\varepsilon$

$\Pi = -B^{-1}\Gamma \qquad \nu = B^{-1}\varepsilon$

$Y = \Pi X + \nu$ - приведенная форма модели.

Проблема идентифицируемости модели.

Структурная форма.

Если выразить Y в каждом уравнении, то экзогенные переменные будут коррелировать с остатками, значит, оценки будут несостоятельными.

Приведенная форма.

Так как переменные X не коррелируют с остатками, то можно применять ОМНК и получать состоятельные оценки.

Проблема: Можно ли использовать оценки приведенной формы для нахождения оценок структурной формы?

Структурный коэффициент называется *идентифицирумым,* если он может быть вычислен на основе коэффициентов приведенной формы.

Уравнение в структурной форме модели *идентифицируемо*, если идентифицируемы все его коэффициенты.

Отсутствие идентифицируемости означает, что существует бесконечно много моделей, совместимых с данными, и это никак не связано с числом наблюдений.

Не хватает факторов, а не количества наблюдений.

Почему коэффициент не может быть вычислен на основе приведенной формы?

Пусть в приведенной форме найдено:

mk - элементов матрицы *П;*

$\frac{m(m+1)}{2}$ - элементов $\|\mathrm{cov}[\nu_i, \nu_j]\|$.

В структурной форме нужно:

$m^2 - m$ - элементов *B* (отнимается число единичных коэффициентов);

mk - элементов *Г*;

$\frac{m(m+1)}{2}$ - элементов $\|\mathrm{cov}[\varepsilon_i, \varepsilon_j]\|$.

В структурной форме число неизвестных на $m^2 - m$ больше, чем в приведенной форме.

Выход видится во введении дополнительных ограничений на коэффициенты структурной формы.

Рассмотрим частный случай.

Условие идентифицируемости.

Необходимо, зная *П*, определить *B* и *Г*.

Рассмотрим одно из уравнений системы в структурной форме (для наглядности – первое).

Предположим, что в этом уравнении равны нулю последние $m - q$, $k - p$ коэффициентов соответственно при *Y* и *X*.

$$(\beta_{11}\ \beta_{12}\ \ldots\beta_{1q}\ 0\ \ldots 0)(Y_1\ Y_2\ldots Y_q Y_{q+1}\ldots Y_m)^T + (\gamma_{11}\ \gamma_{12}\ \ldots\gamma_{1p}\ 0\ \ldots 0)(X_1\ X_2\ldots X_p X_{p+1}\ldots X_m)^T = \varepsilon_1$$

Обозначим

$$\beta_* = (\beta_{11}\ \beta_{12}\ \ldots\beta_{1q}) \qquad Y_* = (Y_1\ Y_2\ldots Y_q)^T \qquad Y_{**} = (Y_{q+1}\ldots Y_m)^T$$

$$\gamma_+ = (\gamma_{11}\ \gamma_{12}\ \ldots\gamma_{1p}) \qquad X_+ = (X_1\ X_2\ldots X_p)^T \qquad X_{++} = (X_{p+1}\ldots X_m)^T$$

Тогда $\beta_* Y_* + \gamma_* X_* = \varepsilon_1$

Рассмотрим приведенную форму.

$$\begin{pmatrix} Y_* \\ Y_{**} \end{pmatrix} = \begin{pmatrix} \Pi_{*,+} & \Pi_{*,++} \\ \Pi_{**,+} & \Pi_{**,++} \end{pmatrix} \begin{pmatrix} X_+ \\ X_{++} \end{pmatrix} + \nu$$

Выразим коэффициенты структурной формы через известные коэффициенты матрицы П.

$$\Pi = -B^{-1}\,\Gamma \qquad \Rightarrow \quad B\Pi = -\Gamma$$

Для первого уравнения: $(\beta_* \quad 0_{m-q}) \begin{pmatrix} \Pi_{*,+} & \Pi_{*,++} \\ \Pi_{**,+} & \Pi_{**,++} \end{pmatrix} = -(\gamma_+ \quad 0_{k-p})$ или

$\beta_* \Pi_{*,+} = -\gamma$

$\beta_* \Pi_{*,++} = 0$

Из последнего выражения имеем *k-p* уравнений с *q*-1 неизвестным.

Необходимое и достаточное условие совместности и определенности системы:

$\operatorname{rang} \Pi_{*,++} = q - 1$ - *ранговое условие* идентифицируемости

Матрица $\Pi_{*,++}$ является расширенной матрицей неоднородной системы, так как выполняется условие нормировки. Ранговое условие означает, что определитель (q-1)-го порядка этой матрицы не равен нулю, т.е. (q-1) строк линейно независимы и (q-1) столбцов линейно независимы, т.е. определитель матрицы системы также равен (q-1). Здесь важнее условие определенности.

Необходимое условие: число уравнений должно быть не меньше числа неизвестных:

$k - p \geq q - 1$ - *порядковое условие* идентифицируемости

Условие не является достаточным, так как в числе уравнений могут быть зависимые, и тогда ранг матрицы системы может быть меньше числа неизвестных (даже при условии совместности), и имеем множество решений – однозначного решения найти нельзя.

Может быть так, что ранг матрицы системы равен рангу расширенной матрицы системы, и число независимых уравнений больше числа неизвестных. При этом порядковое условие выполняется с неравенством – сверхидентифицируемо.

Если порядковое условие выполняется с равенством, то уравнение точно идентифицируемо.

Рассмотрим **пример.** Изучается модель вида

$$\begin{cases} C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C_t – расходы на потребление в период t, Y_t – совокупный доход в период t, I_t – инвестиции в период t, r_t – процентная ставка в период t, M_t – денежная масса в период t, G_t – государственные расходы в период t, C_{t-1} – расходы на потребление в период $t-1$, I_{t-1} инвестиции в период $t-1$. Первое уравнение – функция потребления, второе уравнение – функция инвестиций, третье уравнение – функция денежного рынка, четвертое уравнение – тождество дохода.

Модель представляет собой систему одновременных уравнений. Проверим каждое ее уравнение на идентификацию

Модель включает четыре эндогенные переменные (C_t, I_t, Y_t, r_t) и четыре предопределенные переменные (две экзогенные переменные – M_t и G_t и две лаговые переменные – C_{t-1} и I_{t-1}).

Проверим необходимое условие идентификации для каждого из уравнений модели.

Первое уравнение: $C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1$. Это уравнение содержит две эндогенные переменные C_t и Y_t и одну предопределенную переменную C_{t-1}. Таким образом, $H = 2$, а $D = 4 - 1 = 3$, т.е. выполняется условие $D + 1 > H$. Уравнение сверхидентифицируемо.

Второе уравнение: $I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2$. Оно включает две эндогенные переменные I_t и r_t и одну экзогенную переменную I_{t-1}. Выполняется условие $D + 1 = 3 + 1 > H = 2$. Уравнение сверхидентифицируемо.

Третье уравнение: $r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3$. Оно включает две эндогенные переменные Y_t и r_t и одну экзогенную переменную M_t. Выполняется условие $D + 1 = 3 + 1 > H = 2$. Уравнение сверхидентифицируемо.

Четвертое уравнение: $Y_t = C_t + I_t + G_t$. Оно представляет собой тождество, параметры которого известны. Необходимости в идентификации нет.

Проверим для каждого уравнения достаточное условие идентификации. Для этого составим матрицу коэффициентов при переменных модели.

	C_t	I_t	r_t	Y_t	C_{t-1}	I_{t-1}	M_t	G_t
I уравнение	–1	0	0	b_{11}	b_{12}	0	0	0
II уравнение	0	–1	b_{21}	0	0	b_{22}	0	0
III уравнение	0	0	–1	b_{31}	0	0	b_{32}	0
Тождество	1	1	0	–1	0	0	0	1

В соответствии с достаточным условием идентификации ранг матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, должен быть равен числу эндогенных переменных модели без одного

Первое уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	I_t	r_t	I_{t-1}	M_t	G_t
II уравнение	–1	b_{21}	b_{22}	0	0
III уравнение	0	–1	0	b_{32}	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{22} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{22}b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Второе уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	Y_t	C_{t-1}	M_t	G_t
I уравнение	–1	b_{11}	b_{12}	0	0
III уравнение	0	b_{31}	0	b_{32}	0
Тождество	1	–1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Третье уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	I_t	C_{t-1}	I_{t-1}	G_t
I уравнение	–1	0	b_{12}	0	0
II уравнение	0	–1	0	b_{22}	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{22} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицируемы. Приведенная форма модели в общем виде будет выглядеть следующим образом:

$$1.\ \begin{cases} C_t = A_1 + \delta_{11}C_{t-1} + \delta_{12}I_{t-1} + \delta_{13}M_t + \delta_{14}G_t + u_1, \\ I_t = A_2 + \delta_{21}C_{t-1} + \delta_{22}I_{t-1} + \delta_{23}M_t + \delta_{24}G_t + u_2, \\ r_t = A_3 + \delta_{31}C_{t-1} + \delta_{32}I_{t-1} + \delta_{33}M_t + \delta_{34}G_t + u_3, \\ Y_t = A_4 + \delta_{41}C_{t-1} + \delta_{42}I_{t-1} + \delta_{43}M_t + \delta_{44}G_t + u_1. \end{cases}$$

Оценивание систем одновременных уравнений.

Пусть первое (для определенности) уравнение идентифицируемо.

Косвенный МНК.

2. $y_1 = \Pi X + \nu$

3. $B^{-1}\Gamma = \Pi$

Алгоритм: МНК $\Rightarrow$ $\hat{\Pi}$ $\Rightarrow$ $\hat{B}, \hat{\Gamma}$.

По теореме Слуцкого оценки являются состоятельными, так как они являются непрерывными функциями состоятельных оценок.

Двухшаговый МНК.

4. $BY + \Gamma X = \varepsilon$

5. $B = \|\beta_{ij}\|_{m\times m}$, $\Gamma = \|\gamma_{ij}\|_{m\times k}$, $Y = (Y_1, Y_2, \ldots, Y_m)^T$, $X = (X_1, X_2, \ldots, X_k)^T$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)^T$.

6. $E[\varepsilon] = 0$,

7. $\Sigma[\varepsilon] = \sigma_{ij}I_n$, $(i, j = 1, 2, \ldots, m)$, $\sigma_{ij} = \mathrm{cov}[\varepsilon_{it}, \varepsilon_{js}] = \begin{cases} \neq 0, & t = s \\ 0, & t \neq s \end{cases}$, $(t, s = 1, 2, \ldots, n)$.

8. $\mathrm{cov}[X, \varepsilon] = 0$

9. $|B| \neq 0$

10. Условие нормировки: один из коэффициентов при Y в каждом уравнении равен единице.

Пусть в 1-м уравнении равны нулю последние $m - q$, $k - p$ коэффициентов при *Y* и *X*.

Выразим из этого уравнения одну переменную:

$$Y_1 = \beta_{2-q} Y_{2-q} + \Gamma_{1-p} X_{1-p} = \varepsilon_1$$

Столбцы Y_{2-q} коррелируют с ε_1, так как в другом уравнении Y_{2-q} выразится через Y_1, а, значит, через ε_1. Поэтому оценки МНК не будут состоятельными.

Идея двухшагового МНК: использование X, не входящих в X_{1-p} в качестве ИП для Y_1.

Метод является способом выбора ИП.

Условие идентифицируемости позволяет использовать *n-p* ИП для *q*-1 переменных (условие применения ИП: m>p – число ИП должно быть не меньше числа заменяемых переменных).

Если брать только из X, входящих в X_{1-p}, то будет полная коллинеарность.

Алгоритм:

· Регрессия $Y_1 = X\Pi_1 + \zeta$ - на все экзогенные переменные.

· Регрессия $Y_1 = \beta_{2-q} \hat{Y}_1 + \Gamma_{1-p} X_{1-p} + \varepsilon_1$, где $\hat{Y}_1 = X\hat{\Pi}_1$

Замечания:

· Если $\operatorname{rang} \Pi_{*,++} = q - 1$, $k - p = q - 1$, то оценка косвенного МНК совпадает с оценкой двухшагового метода.

· Оценка двухшагового МНК совпадает с оценкой ИП, если в качестве ИП берут $\hat{Y}_1, X_{1-p}$.

· Если в качестве ИП для Y_1 выбирать любые линейные комбинации из X, то матрица ковариаций оценки будет не меньше матрицы ковариаций двухшагового МНК, т.е. он дает оценку, эффективную в некотором классе оценок.

Фактически оценивается каждое уравнение. Взаимодействие уравнений учитывается путем применения трехшагового МНК.

Вопросы для самопроверки

· Что такое структурная форма системы эконометрических уравнений?
· Что такое приведенная форма системы эконометрических уравнений?
· В чем состоит проблема идентифицируемости системы эконометрических уравнений?
· Критерии идентифицируемости системы эконометрических уравнений?
· Каковы методы идентификации системы эконометрических уравнений?
· Каков алгоритм косвенного МНК?
· Каков алгоритм двухшагового МНК?

Дополнительная литература

· Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 2001. – 430 с. (глава 4).

· Доугерти К. Введение в эконометрику. – М.: ИНФРА – М, 2009. – 465 с. (Глава 9).

· Елисеева И.И. Эконометрика: Учебник / под. ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 344 с. (глава 5).

Интернет-ресурсы

· http://www.nsu.ru/ef/tsy/ecmr/index.htm
· http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html
· http://www.statsoft.ru/home/textbook/glossary/default.htm

Тема 6. Модели временных рядов

Элементы временного ряда. Идентификация структуры временного ряда. Автокорреляционная и частная автокорреляционная функции. Аддитивная и мультипликативная модели. Модели Бокса-Дженкинса.

Задачи изучения темы:

· научиться выделять основные компоненты временного ряда,

· научиться определять автокорреляционную функцию временного ряда и использовать ее для исследования ряда.

· научиться строить мультипликативную и аддитивную модели временного ряда и использовать их для прогнозирования,

· получить представление о методологии исследования временных рядов Бокса-Дженкинса.

Теоретический материал

Компоненты уровней ряда динамики.

· *Основная тенденция развития (тренд)* (T) – результат влияния постоянно действующих факторов.

· *Сезонная составляющая* (S) – результат влияния периодически действующих факторов.

· *Случайная компонента (ошибка)* (E) – результат влияния случайных факторов.

Общая модель временных рядов.

Уровень ряда Y представляет собой функцию от указанных компонент:

$$Y = f(T, S, E).$$

Виды моделей временных рядов.

В зависимости от вида функции различают следующие виды моделей временных рядов:

· *Аддитивная модель*:

$$Y = T + S + E;$$

· *Мультипликативная модель:*

$$Y = T \cdot S \cdot E.$$

Каждому виду модели соответствует свои методы определения составляющих компонент.

Изучение тренда.

Виды трендов.

· *Тенденция среднего уровня* – детерминированная составляющая явления

· *Тенденция дисперсии* – характеризует динамику отклонений между эмпирическими уровнями и детерминированной компонентой ряда

· *Тенденция автокорреляции* – характеризует тенденцию изменения связи между отдельными уровнями ряда динамики.

Этапы изучения тренда.

· Тестирование ряда динамики на наличие тренда

· Выделение тренда (выравнивание временного ряда).

Проверка на наличие тренда.

· *Метод средних:*

○ Ряд разбивается на интервалы (обычно на два).

○ Проверяется гипотеза о равенстве средних: $H_0: \quad \bar{y}_1 = \bar{y}_2$

· *Критерий Кокса и Стюарта:*

○ Ряд разбивают на три части

○ Сравнивают средние уровни крайних групп.

· *Фазочастотный критерий знаков первой разности* (Валлиса и Мура):

○ Вычисляют абсолютные цепные приросты

○ Наличие тренда утверждается, если ряд не содержит или мало содержит фазы – изменение знака абсолютных цепных приростов (первых разностей).

· *Метод серий*:

○ Уровни относят к двум классам в зависимости от сравнения с медианой.

○ Определяется число серий – последовательностей уровней одного класса. Число серий – нормально с параметрами: $(n+1)/2$ и $\sqrt{(n-1)/4}$, если тенденции нет.

○ Рассчитывается доверительный интервал с заданной вероятностью.

○ Тенденция считается существующей, если число серий выходит за пределы этого интервала.

Методы выделения тренда.

· *Метод укрупнения интервалов.*

Определяются значения уровней по укрупненным периодам времени

· *Метод скользящей средней.*

Вычисляется последовательность значений средних уровней из определенного числа уровней ряда, начиная с первого, второго, третьего и т.д.

· *Метод аналитического выравнивания*.

Определяется функциональная зависимость значений уровней ряда от времени.

○ Метод конечных разностей:

$$\bar{y}_t = y_0 + t\Delta_0^{(1)} + \frac{t(t-1)}{2!}\Delta_0^{(2)} + \frac{t(t-1)(t-2)}{3!}\Delta_0^{(3)} + \ldots + \Delta_0^{(t)},$$

где $\Delta_t^{(1)} = y_{t+1} - y_t$, $\Delta_t^{(2)} = \Delta_{t+1}^{(1)} - \Delta_t^{(1)}$ и т.д. – конечные разности.

○ Метод наименьших квадратов с использованием функций вида:

$$\bar{y}_t = a_0 + \sum_{i=1}^{n} a_i t^i, \qquad \text{(в частности, } \hat{y}(x) = \hat{a}t + \hat{b}\text{) или}$$

$$\bar{y}_t = a_0 + \sum_{i=1}^{n} (a_i \cos it + b_i \sin it).$$

Использование модели тренда для оценки случайной компоненты

Колеблемость уровней около тренда служит мерой воздействия остаточных факторов.

Для ее измерения используются следующие показатели

Стандартная ошибка.

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_{it})^2}{n}}$$

Коэффициент вариации.

$$v = \frac{\sigma_t}{\bar{y}}$$

Анализ сезонных колебаний.

Модель сезонной волны.

Модель сезонной волны – система индексов сезонности $I_1, I_2, \ldots, I_k$ (k - число сезонов в году).

Индексы сезонности.

Индексы сезонности – отношения фактических внутригодовых уровней к среднему уровню.

Методы определения индексов сезонности.

· *Метод постоянной средней* – в случае отсутствия тренда среднего уровня:

$$I_i = \frac{y_i}{\bar{y}} 100\% \quad \text{или} \quad I_i = \frac{\bar{y}_i}{\bar{y}} 100\%, \quad i = 1,2,...k.$$

$\bar{y}_i$ - среднее значение уровней по одноименному сезону i, рассчитанное за несколько лет.

· *Метод переменной средней* – учетом основной тенденции развития среднего уровня:

$$I_i = \frac{y_i}{\bar{y}_{ti}} 100\% \quad \text{или} \quad I_i = (\sum_{j=1}^{n} \frac{y_j}{\bar{y}_{tj}}) : n, \quad i = 1,2,...k.$$

n - число лет при усреднении по одноименным сезонам, $\bar{y}_{tj}$ - значение переменной средней, соответствующее
i-му сезону в j-м году.

Построение моделей временных рядов.

Алгоритм построения модели.

· Выравнивание исходного ряда методом скользящего среднего.

· Расчет значений сезонной компоненты S.

· Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных ($T*E$ или $T + E$).

· Аналитическое выравнивание уровней $T*E$ или $T + E$ и расчет значений T с использованием полученного уравнения тренда.

· Расчет полученных по модели значений $T*E$ или $T + E$.

· Расчет абсолютных и/ или относительных ошибок.

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

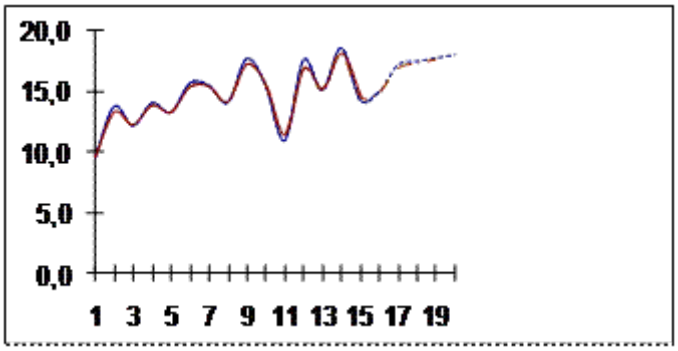
· Использование моделей для прогнозирования.

· Сравнение моделей: (аддитивной, мультипликативной, модели экспоненциального сглаживания).

Мультипликативная модель.

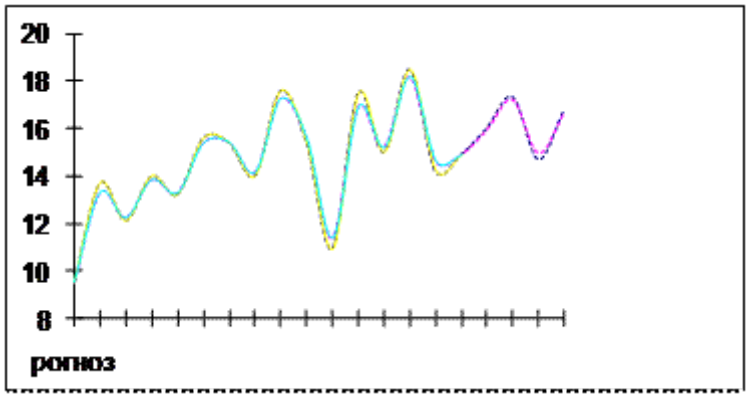
$$Y = T*S*E$$

Для мультипликативной модели характерна относительно высокая изменчивость частоты колебаний значений исследуемого показателя.



Аддитивная модель.

$$Y = T + S + E$$



Автокорреляция уровней ряда и выявление его структуры

Корреляционную зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда***.**

Автокорреляция уровней ряда измеряется с помощью *коэффициента автокорреляции*.

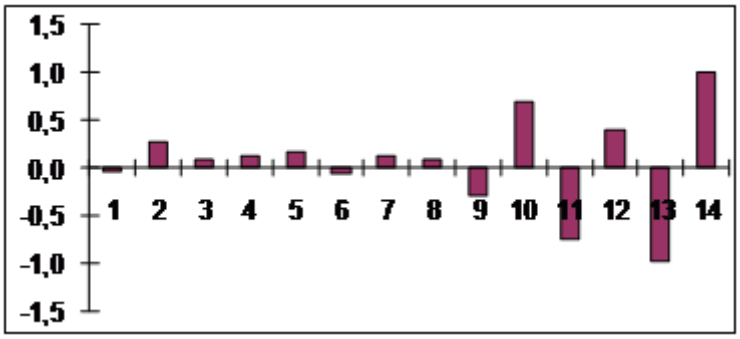
Число периодов, по которым рассчитывается коэффициент автокорреляции, назывется *лагом.*

Последовательность коэффициентов автокорреляции называется *автокорреляционной функцией временного ряда.*

Свойства автокорреляционной функции.

- Характеризует тесноту линейной связи, поэтому можно судить о наличии линейной тенденции, он может быть равен нулю для нелинейной тенденции.
- По знаку нельзя судить о направлении монотонности тенденции.

График зависимости корреляционной функции от величины лага называется коррелограммой.



Анализ автокорреляционной функции позволяет выделить лаг, при котором связь между текущим и предыдущими уровнями наиболее тесная, т.е. можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент первого порядка, то ряд содержит только тенденцию

Если наиболее высоким оказался коэффициент k-го порядка, то ряд содержит циклические (сезонные) колебания с периодом k.

Если ни один из коэффициентов не является значимым, то ряд либо не содержит тенденции и циклических колебаний и имеет структуру случайной компоненты, либо ряд содержит сильную нелинейную тенденцию, для выявления которой требуется дополнительный анализ.

Поэтому автокорреляционная функция широко используется при анализе временных рядов для выявления тренда и циклических (сезонных) колебаний.

Прогнозирование.

Основой распространения тенденции на будущее является свойство инерционности социально-экономических явлений, состоящее в том, что закономерность развития, действующая в прошлом, сохранится и в прогнозируемом будущем, т.е. прогноз базируется на перспективной экстраполяции.

Чем короче срок экстраполяции, тем более надежны и точны результаты прогнозирования.

В общем виде экстраполяцию можно представить в виде функции вида:

$$\hat{y}_{i+T} = f(y_i, T, a_j),$$

где $\hat{y}_{i+T}$ - прогнозируемый уровень, y_i - текущий уровень прогнозируемого ряда, T - период прогноза, a_j - параметр уравнения тренда.

Методы экстраполяции.

· *Метод среднего абсолютного прироста* – при условии стабильности абсолютных приростов:

$$\hat{y}_{i+T} = y_i + T\bar{\Delta},$$

$\bar{\Delta}$ – средний абсолютный прирост.

· *Метод среднего темпа роста* – при условии развития изучаемого явления по экспоненте:

$$\hat{y}_{i+T} = y_n \cdot \overline{Tp},$$

$\overline{Tp}$ – средний темп роста, y_n – последний уровень ряда.

· *Метод использования аналитического выравнивания* – прогнозируемое значение получают с учетом модели тренда при $t = T$.

При этом в качестве интервальной оценки прогноза используют доверительный интервал в виде:

$$\hat{y}_t \pm t_\gamma \cdot \sigma_{\bar{y}_t},$$

где $\hat{y}_t$ - расчетное значение прогнозируемого уровня, $\sigma_{\bar{y}_t} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ti})^2}$ - средняя квадратическая ошибка тренда, t_γ - табличное значение распределения Стьюдента при заданной доверительной вероятности γ.

Вопросы для самопроверки

· Чем отличаются модели временных рядов от регрессионных моделей?
· Каковы основные виды моделей временных рядов?
· В чем состоит суть компонентного анализа временных рядов?
· Каков алгоритм построения аддитивной модели временного ряда?
· Каков алгоритм построения мультипликативной модели временного ряда?
· Каков алгоритм исследования тренда?
· Каков алгоритм исследования сезонной компоненты?
· Каковы методы изучения случайной компоненты?
· Что такое автокорреляционная функция временного ряда?
· Каков алгоритм определения построения автокорреляционной функции данного временного ряда?
· Какова суть методологии Бокса-Дженкинса?

Дополнительная литература

· Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 2001. – 430 с. (глава 3).
· Доугерти К. Введение в эконометрику. – М.: ИНФРА – М, 2009. – 465 с. (Глава 3).
· Елисеева И.И. Эконометрика: Учебник / под. ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 344 с. (глава 6-15).

Интернет-ресурсы

· http://www.nsu.ru/ef/tsy/ecmr/index.htm
· http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html

· http://www.statsoft.ru/home/textbook/glossary/default.htm

Дополнительная литература по дисциплине

· Айвазян С.А. Иванова С.С. Эконометрика. Краткий курс: учеб. пособие. – М.: Маркет ДС, 2007. – 104 с.
· Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 2001. – 430 с. (глава 1, глава 2, п.2.1).
· Доугерти К. Введение в эконометрику. – М.: ИНФРА – М, 2009. – 465 с.
· Елисеева И.И. Эконометрика: Учебник / под. ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 344 с.

Интернет-ресурсы

· http://www.nsu.ru/ef/tsy/ecmr/index.htm
· http://subscribe.ru/archive/science.humanity.econometrika/200007/17050500.html
· http://www.statsoft.ru/home/textbook/glossary/default.htm