НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

1930

МЭИ

Курс лекций по дисциплине: Эконометрика

ФИО: Скиба Мария Сергеевна,
сотрудник кафедры ЭЭП НИУ «МЭИ»
Почта: SkibaMS@mpei.ru

Москва, 2021**г.**

Темы курса:

- Введение в эконометрику
- Регрессия
- Парная нелинейная регрессия
- Множественная регрессия корреляция
- Мультиколлинеарность
- Система эконометрических уравнений

Введение в эконометрику

Эконометрика -наука, предметом которой является количественное выражение взаимосвязей экономических явлений и процессов.

Цель: Разработка способов моделирования количественного анализа реальных экономических объектов.

Задачи эконометрики:

- Спецификация моделей
- Параметризация моделей
- Верификация моделей
- Прогнозирование моделей

Эконометрическая модель в общем виде : $y = f(x) + \varepsilon$

У- результативный признак

Х- факторный признак

Е- Случайная переменная

Задачи эконометрического моделирования:

- Определить объяснимую часть, то есть найти У
- Получить оценки параметров случайной составляющей, то есть найти Е

Классы эконометрических моделей :

1. Регрессионные модели с одним уравнением $y_1=f(x_1,x_2 \dots x_n)$
2. Системы одновременных уравнений $y_2=f(x_1,x_2 \dots x_n,y_1,y_2 \dots y_n)$
3. Модели временных рядов $y=f(t)$

Типы данных:

1. Пространственные
2. Временные

Виды переменных:

Экзогенные → Эндогенные → Лаговые

Этапы эконометрического моделирования:

- ❖ Постановочный
- ❖ Априорный
- ❖ Параметризация
- ❖ Информационный
- ❖ Идентификация
- ❖ Верификация

Виды зависимостей :

Функциональная

Статическая

Корреляционная

Классификация корреляционной зависимости:

1. По направлению действия (прямая и обратная)
2. По аналитическому выражению (прямолинейная и криволинейная)
3. По количеству признаков (однофакторные и многофакторные)

Что такое корреляция и корреляционный анализ?

Виды корреляции: 1) Парная
2) Множественная
3) Частная

Коэффициент корреляции

$$ryx=\frac{cov(y*x)}{\partial x*\partial y}$$ ∂ – сигма

$cov(yx)=\overline{yx}-\overline{y}*\overline{x}$ $\overline{*}$ - среднее значение

$$\partial x=\sqrt{\frac{\sum_{i=1}^{n}(xi-\overline{x})^2}{n}} \qquad \partial y=\sqrt{\frac{\sum_{i=1}^{n}(yi-\overline{y})^2}{n}}$$

Если коэффициент корреляции=0 , то связь между признаками ***отсутствует***

Если от +- 0 до +- 0,3 , то связь между признаки ***практически отсутствует***

Если от +- 0,3 до +- 0,5 то связь между признаками ***слабая***

Если от +- 0,5 до +- 0,7 , то связь между признаками ***умеренная***

Если от +- 0,7 до +- 1 , то связь ***сильная***

Cov показывает направление связи : если >0 ,то связь прямая. Если <0, то связь ***обратная***

Этапы корреляционно-регрессионного анализа:

1) Предварительный анализ явлений и выявление причин возникающих взаимосвязей между признаками;
2) Разделение признаков на факторные и результативные;
3) Построение матрицы парной корреляции;
4) Оценка формы уравнения регрессии;
5) Решение уравнения регрессии;
6) Расчет теоретического ожидания значения результативного признака;
7) Определение и сравнение анализа дисперсии;
8) Оценка тесноты связей между признаками;
9) Общая оценка качества модели;
0) Статистическая оценка достоверности параметров уравнения регрессии;
1) Построение доверительных границ для теоретического ожидания значения функции;
2) Формулирование вывода из анализа;

Регрессия

Регрессия -уравнение связей нескольких переменных

Цель: Оценка функциональной зависимости условного среднего значения

Результативного признака от факторного признака

Регрессионный анализ заключается в определении аналитической формы связей, в которой изменение результативного признака обусловлено влиянием факторного признака.

Виды регрессии:

1) Парная – связь между результативным и факторным признаками $\hat{y} = \mathrm{f}(x)$
2) Множественная – связь между результативным и несколькими факторными признаками $\hat{y} = \mathrm{f}(x_1, x_2 \dots x_n)$

Парная регрессия: Линейная $\hat{y} = a + bx$

Система нормальных уравнений :

$$\begin{cases} \mathrm{na} + b * \sum_{i=1}^{n} \mathrm{xi} = \sum_{i=1}^{n} yi \\ a * \sum_{i=1}^{n} \mathrm{xi} + b * \sum_{i=1}^{n} xi^2 = \sum_{i=1}^{n} xi * yi \end{cases}$$

$$\mathbf{b} = \frac{\mathrm{cov}(\mathrm{xy})}{\mathrm{cov}(x_1 \mathrm{x})} = \frac{\overline{yx} - \overline{y} * \overline{x}}{\overline{x^2} - (\overline{x})^2}$$

$$\mathrm{a} = \overline{y} - b * \overline{x}$$

Коэффициенты, характеризующие уравнения регрессии

Коэффициент эластичности $Э = f'(x) * \frac{\bar{x}}{\bar{y}}$

Бета коэффициент $\beta = b * \frac{\partial x}{\partial y}$

Коэффициент детерминации $R = r^2 yx$

Проверка адекватности и точности уравнений регрессии:

Этапы:

1) Определение значимости модели
2) Оценка качества параметра модели
3) Установление систематической ошибки

Значения модели определяются с помощью критерия Фишера $\frac{r_{yx}^2 * (n-m-1)}{(1-r_{yx}^2)*1}$

если при заданном уровне значимости $\alpha = 0{,}05$ рассчитанное значение Fкр Фишера получилось больше табличного , то модель считается значимой и отклоняется гипотеза о случайной природе оцениваемых характеристик. F>таб.значения

Качество параметров модели определяется с помощью t кр стьюдента

$$m_a = \frac{\partial_{\text{ост}} * \sqrt{\sum_{i=1}^{n} xi^2}}{n * \partial_x} \qquad m_b = \frac{\partial_{\text{ост}}}{\partial_x * \partial_y} \qquad m_r = \frac{\sqrt{\sum_{i=1}^{n} (yi - \widehat{yi})^2}}{n}$$

$$t_a = \frac{a}{m_a} \qquad t_b = \frac{b}{m_b} \qquad t_r = \frac{r_{yx}}{m_r}$$

Если рассчитанное значение t критерия Стьюдента получилось больше табличного , то параметр считается качественным и статистически значимым

Для таблицы : Df=n-m-1 $\alpha = 0,05$

Расчет доверительных интервалов

Для расчета доверительных интервалов находят предельную ошибку

$$\Delta a = a * t_{\text{табл}} \qquad \Delta b = b * t_{\text{табл}}$$

$$\gamma_{min}^{a} = a - \Delta a \qquad \gamma_{max}^{a} = a + \Delta a$$

$$\gamma_{min}^{b} = b - \Delta b \qquad \gamma_{max}^{b} = b + \Delta b$$

Определение меры точности модели :

1) Максимальная ошибка

$$\epsilon^{max} = y - \hat{y}$$

2) Средняя абсолютная ошибка

$$\overline{\varepsilon}^{\text{абсолютная}} = \frac{\sum_{i=1}^{n}|\varepsilon|}{n}$$

3) Дисперсия ряда остатков

$$S^2 = \frac{\sum_{i=1}^{n}(\varepsilon - \overline{\varepsilon})^2}{n-1}$$

4) Среднеквадратическое отклонение

$$\partial_{\text{ост}} = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}}$$

5) Средняя ошибка аппроксимации

$$\overline{A} = \left|\frac{\frac{y-\hat{y}}{y}}{n}\right| * 100\%$$

Прогнозирование: Прогноз

1) Точечный
2) интервальный

$$X_{\text{пр}} = \hat{y}\text{пр} \quad \hat{y}\text{пр} = a + b * X_{\text{пр}} \quad m\hat{y}\text{пр} = \partial_{\text{ост}} * \sqrt{1 + \frac{1}{n} + \frac{(y\text{пр} - \bar{x})^2}{\sum_{i=1}^{n}(xi - \bar{x})^2}}$$

$$\Delta\hat{y}\text{пр} = m\hat{y}\text{пр} * t\text{табл} \quad \gamma_{min}^{\hat{y}\text{пр}} = \hat{y}\text{пр} - \Delta\hat{y}\text{пр} \quad \gamma_{max}^{\hat{y}\text{пр}} = \hat{y}\text{пр} + \Delta\hat{y}\text{пр}$$

Требования к построению уравнения регрессии:

1) Совокупность исходных данных должна быть однородной и математически описываться непрерывными функциями
2) Необходимо наличие достаточно большого объема исследуемой выборочной совокупности
3) Моделированные явления должны описываться одним или несколькими причинно-следственная связями
4) Причинно-следственные связи по возможности следует описывать линейной или приводимой к линейной форме зависимости
5) Должны отсутствовать количественные ограничения на параметры
6) Факторные признаки должны выражаться количественно
7) Необходимо постоянно территориальной и временной структуры изучаемой совокупности

Условия теоретической обоснованности моделей:

1) Все признаки и их совместные распределения должны подчиняться нормальному закону
2) Дисперсия моделируемого признака должна оставаться постоянной при изменении факторных признаков

Дисперсия -отклонение от математического ожидания

Математическое ожидание – среднее значение вероятностей

3) Отдельные наблюдения должны быть зависимыми:

Требования ,при которых модель считается адекватной:

1) Уровни ряда остатков имеют случайный характер . Проверяется с помощью критерия поворотных точек
2) Математическое ожидание уровней ряда остатков=0
3) Дисперсия каждого отклонения одинакова для всех значений x

- если дисперсия одинаковая , то это гомоскедастичность
- если дисперсия разная ,то это гетероскедастичность (используется метод Гольдфельда –Кванта)

1) Отсутствует автокорреляция, т.е значение уровней ряда остатков независимых друг от друга
2) Уровни ряда остатков распределены по нормальному закону

Проверяется с помощью R/S критерия

Парная нелинейная регрессия:

Классы нелинейной регрессии:

1) регрессии нелинейные по переменным ,но линейны по оцениваемым параметрам(гипербола , полинома)
2) регрессии нелинейные по переменным и по оцениваемым параметрам (показательная, степенная , экспоненциальная)

Подходы оценки параметров нелинейных моделей:

1) линеаризация
2) метод нелинейной оптимизации

Гипербола

$\hat{y} = a + b * \frac{1}{x}$ $\frac{1}{x} = z$ $\hat{y} = a + b * z$ $b = \frac{\overline{yz} - \overline{y} * \overline{z}}{\overline{z^2} - \overline{z}^2}$ $a = \overline{y} - b * \overline{z}$

Степенная

$\hat{y} = a * x^b$ $lg\hat{y} = lga + b * lgx$ lgy=Y lga=C lgx=X

$\hat{y} = c + b * x$ $b = \frac{\overline{yx} - \overline{y} * \overline{x}}{\overline{x^2} - \overline{x}^2}$ $c = \overline{y} - b * \overline{x}$ a=10^c

Показательная

$\hat{y} = a * b^x$ $lg\hat{y} = lga + x * lgb$ lgy=Y lga=C lgb=B

$B = \frac{\overline{yx} - \overline{y} * \overline{x}}{\overline{x^2} - \overline{x}^2}$ $C = \overline{Y} - B * \overline{x}$ a=10^c

Теснота связей

$$\rho yx = \sqrt{1 - \frac{\partial^2_{ост}}{\partial^2_y}}$$

ρ =0, то связь между признаками <u>отсутствует</u>

Если от 0 до 0,3, то связь между признаки <u>практически отсутствует</u>

Если от 0,3 до 0,5, то связь между признаками <u>слабая</u>

Если от 0,5 до 0,7, то связь между признаками <u>умеренная</u>

Если от 0,7 до 1, то связь <u>сильная</u>

Дисперсионный анализ:

Любая сумма квадратов отклонений связана с числом степеней свободы(df)

Суммы отклонений :

1) общая $\sum_{i=1}^{n}(y-\bar{y})^2$

   Sобщ= Sост+Sфакт

   Sdf= dfост+dfфакт

1) Остаточная $\sum_{i=1}^{n}(y-\hat{y})^2$
2) Факторная $\sum_{i=1}^{n}(\hat{y}-\bar{y})^2$

вариация	Число степ. свободы df	Сумма квадратов s	Дисперсия на одну степ. свободы D	F	Fтабл
общая	n-1			-	-
факторная	m	Sобщ- Sост			табл
Остаточная	n-m-1			-	-

Множественная регрессия и корреляция:

Задачи множественно-корреляционно-регрессионного анализа:

1. Изменение тесноты связи между признаками
2. Отбор факторных признаков в модели
3. Установление неизвестных причин
4. Определение вида уравнении регрессии
5. Построение регрессионной модели и оценка ее параметрах
6. Проверка значимости параметров связи
7. Интервальное оценивание параметров связи

$\hat{y} = f(x1, x2, x3 \ldots xn)$

$\hat{y} = a + b1x1 + b2x2$ *-уравнение регрессии в естественной форме*

$ty = B1tx1 + B2tx2$ *-стандартизированный масштаб*

$ty = \frac{y_i - \bar{y}}{\bar{y}}$

$tx = \frac{x_i - \bar{x}}{\bar{x}}$

Матричная форма:

$\hat{y} = \mathrm{X} * \mathrm{A}$

$$У = \begin{bmatrix} y1 \\ y2 \\ yn \end{bmatrix}$$

$$A = \begin{bmatrix} a \\ b1 \\ b2 \\ bk \end{bmatrix}$$

$$\mathrm{X} = \begin{bmatrix} 1 & \mathrm{X}11 & \mathrm{X}12 & \mathrm{X}1к \\ 1 & X21 & X22 & X2k \\ 1 & Xn1 & Xn2 & Xnk \end{bmatrix}$$

*У-векторное значение зависимой переменной с размерностью (n*1)*

*А-вектор подлежащий оцениванию неизвестных параметров с размерностью (к+1)*1*

Х- Матрица значений независимых переменных (nxk+1)

*1-ый столбец матрицы-единичный так как в уравнении регрессии a*1*

$\hat{y} = \mathrm{a} + \frac{b_1}{x_1} + \frac{b_2}{x_2}$ *Гипербола*

$\hat{y} = \mathrm{a} + \mathrm{x1}^{b1} * \mathrm{x2}^{b2}$ *Степенная*

$\hat{y} = \mathrm{a} * b1^{x1} * b2^{x2}$ *Показательная*

Параметры уравнения регрессии можно найти с помощью методов наименьших квадратов.

Необходимо решить матричное уравнение:

$A=(X'X)^{-1}=X'Y$

X'-транспонированная матрица X

$(X'X)^{-1}$- обратная матрица X(если все это решить, то получим формулы для параметров управления регрессии)

$$b1=\frac{\partial y}{\partial x1}*\frac{r_{yx1}-r_{yx2}*r_{x1x2}}{1-r_{x1x2}^{2}}$$

$$b2=\frac{\partial y}{\partial x2}*\frac{r_{yx2}-r_{yx1}*r_{x1x2}}{1-r_{x1x2}^{2}}$$

$$a=\overline{y}-b1*\overline{x1}-b2*\overline{x2}$$

<u>*Парные коэффициенты корреляции*</u> характеризуют тесноту связи между двумя признаками

<u>*Частные коэффициенты корреляции*</u> характеризуют тесноту связи между 2-мя признаками при фиксированном значении остальных факторных признаков

Парная корреляция остается такой же

$$r_{yx1}=\frac{\overline{yx1}-\overline{y}*\overline{x1}}{\partial y*\partial x1}$$

$$r_{yx2}=\frac{\overline{yx2}-\overline{y}*\overline{x2}}{\partial y*\partial x2}$$

$$r_{x1x2}=\frac{\overline{x1x2}-\overline{x1}*\overline{x2}}{\partial x1*\partial x2}$$

$$r_{yx1x2}=\frac{r_{yx1}-r_{yx2}*r_{x1x2}}{\sqrt{(1-r_{yx2}^{2})*(1-r_{x1x2}^{2})}}$$

$$r_{yx2x1}=\frac{r_{yx2}-r_{yx1}*r_{x1x2}}{\sqrt{(1-r_{yx1}^{2})*(1-r_{x1x2}^{2})}}$$

$$r_{x1x2y}=\frac{r_{x1x2}-r_{yx1}*r_{yx2}}{\sqrt{(1-r_{yx1}^{2})*(1-r_{yx2}^{2})}}$$

Множественный коэффициент корреляции – характеризует тесноту связи между всеми признаками

$$R = \sqrt{1 - \frac{\partial \text{ост}^2}{\partial y^2}}$$

$$R = \sqrt{\sum_{i=1}^{n} B_i * r_{yx1}}$$

$$R = \sqrt{\frac{r_{yx1}^2 + r_{x2}^2 - 2r_{yx1} * ryx2 * rx1x2}{1 - r_{x1x2}^2}}$$

$$R = \sqrt{1 - \frac{\Delta r}{\Delta r11}}$$

$$\Delta r11 = \begin{bmatrix} 1 & rx1x2 \\ rx1x2 & 1 \end{bmatrix}$$

$$\Delta r = \begin{vmatrix} 1 & ryx1 & ryx2 \\ ryx1 & 1 & rx1x2 \\ ryx2 & rx1x2 & 1 \end{vmatrix}$$

Δr – определитель матрицы парный коэффициент корреляции

Множественный коэффициент корреляции может принимать значения от 0 до 1

Коэффициент эластичности остается такой же $Э = f'(x) * \frac{\bar{x}}{\bar{y}}$

Для линейного уравнения $Э = bi * \frac{\bar{xi}}{\bar{yi}}$

Коэффициент детерминации: $R^2 = (R)^2$

Скорректированный коэффициент детерминации: $\hat{R}^2 = 1 - (1 - R^2) * \frac{n-1}{n-m-1}$

Для оценки качества используется критерий Фишера:

Общий: $\boldsymbol{F} = \frac{R^2*(n-m-1)}{(1-R^2)*m}$

Частный: $\boldsymbol{Fx1} = \frac{(R^2-r_{yx2}^2)*(n-m-1)}{1-R^2}$

$$\boldsymbol{Fx2} = \frac{(R^2-r_{yx1}^2)*(n-m-1)}{1-R^2}$$

Для оценки значимости используется критерий Стьюдента

В множественной регрессии считается, что параметр a статистически значим, следовательно, tкр находится только для параметра b

$\boldsymbol{tb1} = \frac{b1}{mb1}$ $\boldsymbol{tb2} = \frac{b2}{mb2}$

Случайная ошибка:

$$mb1 = \frac{\partial y * \sqrt{1 - R^2}}{\partial x1 * \sqrt{1 - r_{x1x2}^2} * \sqrt{n - m - 1}}$$

$$mb2 = \frac{\partial y * \sqrt{1 - R^2}}{\partial x2 * \sqrt{1 - r_{x1x2}^2} * \sqrt{n - m - 1}}$$

Отбор факторных признаков в модель:

Стадии отбора:

1) Предварительное определение перечня факторов, оказывающих влияния на переменную у
2) Сравнительная оценка и отсев факторов
3) Окончательный отбор факторов в процессе построения разных вариантов моделей и оценки значимости их параметров

Для сравнительной оценки и отсева составляется матрица парных коэффициентов корреляции, измеряемая тесноту линейных связей с результативным признаком и каждым фактором- матрица должна быть симметричной.

Матрица позволяет выявлять факторы, которые находятся в тесной линейной корреляционной взаимосвязи близкой к функциональной

	y	X1	X2	…	Xj	…	Xn
y	1	ryx1	ryx2	…	ryxj	…	ryxn
X1	rx1y	1	rx1x2	…	rx1xj	…	rx1xn
X2	rx2y	rx2x1	1	…	rx2xj	…	rx2xn
…	…	…	…	….	….	…	…
xi	rxiy	rxix1	rxix2	…	rx1xj	…	rxixn
…	…	…	…	….	….	…	…
xn	rxny	rxnx1	rxnx2	…	rxnxj	…	1

Мультиколлинеарность:

Мультиколлинеарность- тесная взаимосвязь линейного характера между факторными признаками включенными в модель

Изменения, возникшие под воздействием мультиколлинеарности

1. Искажение величины параметров модели, которые имеют тенденцию к завышению
2. Приводит к изменению смысла экономических интерпретаций параметров регрессии
3. Вызывает слабую обусловленность системы нормальных уравнений
4. Осложняет процесс определения наиболее существенный факторных признаков

Этапы решения проблемы мультиколлинеарности

Искажение величины параметров

Определение причин возникновения

Разработка мер по устранению

Причины возникновения мультиколлинеарности:

1. Изучаемые факторные признаки характеризуют одни и те же сторонние явления или процессы
2. Использование в качестве факторных признаков величин суммарные значения, которых представляет постоянное число
3. Факторные признаки являются элементами друг друга
4. Факторные признаки дублируют друг друга по экономическому смыслу

Методы устранения мультиколлинеарности:

- ❖ Сравнение значений линейных коэффициентов корреляции
- ❖ Метод включения факторов в модель
- ❖ Метод исключения факторов

Система эконометрических уравнений :

Виды систем уравнений:

1. Система независимых уравнений
2. Система рекурсивных уравнений
3. Система одновременных уравнений

Причины, по которым строят приведенную форму модели:

1. Оценки параметров структурной формы являются смещенными и не состоятельными, так как эндогенные переменные коррелируют со случайными отклонениями
2. Уравнения в приведенной форме можно решать с помощью простейшего метода наименьших квадратов
3. Параметры приведенной формы выражаются через параметры структурной формы

Переход от приведенной формы к структурной связан с решением проблемы идентификации

Идентификация – установление соответствия между приведенной и структурной формой модели

Классы моделей с точки зрения идентификации:

1. Идентифицируемые
2. Неидентифицируемые
3. Сверхидентифицируемые

Необходимое условие идентификации:

Д+1=Н

Д+1<Н

Д+1>Н

Методы определения идентификации:

1. Косвенные методы
2. Двухшаговые методы

1. Система независимых уравнений

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ y_2 = a_{21} \cdot x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \ldots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{cases}$$

2. Система рекурсивных уравнений

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12}x_2 + a_{1n}x_n \\ y_2 = a_{21} \cdot x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_{21}y_1 \\ \ldots \\ y_m = a_{m1} x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + b_{m1}y_1 + b_{m2}y_2 + \ldots + b_{m-1} \cdot y_{n-1} \end{cases}$$

3. Система одновременных уравнений

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12} x_2 + \ldots + a_{1n}x_n + b_{12}y_2 + \ldots + b_{1n}y_n \\ y_2 = a_{21} x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_{21}y_1 + \ldots + b_{2n}y_n \\ \ldots \\ y_m = a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn}x_n + b_{m1}y_1 + b_{m2}y_2 + \ldots + b_{m-1} \cdot y_{n-1} \end{cases}$$

Уравнения, записанные в структурной форме нужно перевести в приведенную.

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