

Эконометрика

Полковников Александр Александрович

Волжский политехнический институт (филиал)
ФГБОУ ВПО "Волгоградский государственный технический университет"

Конспект лекций для студентов направления
"Экономика"

Введение

В качестве самостоятельного научного направления эконометрика выделилась в конце 20-х годов XX в.

В 1930 г. было образовано “Эконометрическое общество”. С 1933 г. начал издаваться журнал “Econometrica”.

Определение

*С современных позиций **эконометрику** можно определить как науку о моделировании экономических явлений, позволяющем объяснить и прогнозировать их развитие, выявлять и измерять определяющие факторы.*

Новое знание в экономике может быть получено только посредством математического моделирования. Модель отражает попытку реконструировать в упрощенной форме тот механизм, который скрыт за изучаемыми явлениями. На основе модели формируется предварительное объяснение, которое затем должно пройти проверку путем статистического тестирования.

Построение любой эконометрической модели проходит следующие этапы:

- первый** — теоретический, в ходе которого формулируется цель исследования, определяется круг участвующих в модели экономических характеристик, создается априорное описание формализованных связей между ними;
- второй** — информационный, когда осуществляется поиск требуемых данных, проверяется их достоверность, сопоставимость, осуществляются необходимые пересчёты, используются пространственные и временные данные;
- третий** — спецификация модели, когда устанавливаются экзогенные (внешние) и эндогенные (внутренние) переменные, выявляются связи и соотношения;

- четвертый** — идентификация модели, т. е. выявление условий корректного оценивания параметров модели на основе соотношения количества переменных и связей между ними;
- пятый** — оценка параметров модели;
- шестой** — верификация модели, т. е. проверка адекватности модели, делается вывод о том, какова точность, вычисленных на основе модели, прогнозных оценок; производится анализ остатков (случайных компонент).

§1. Парная линейная регрессия

Определение

Парная регрессия — это уравнение, описывающее корреляционную связь между парой переменных: зависимой переменной (результатом) y и независимой переменной (фактором) x :

$$y = f(x).$$

*Если функция $f(x) = \alpha + \beta x$ — линейная, то говорят о **парной линейной регрессии**.*

Определение

Обычно используют следующее представление:

$$y = f(x) + e.$$

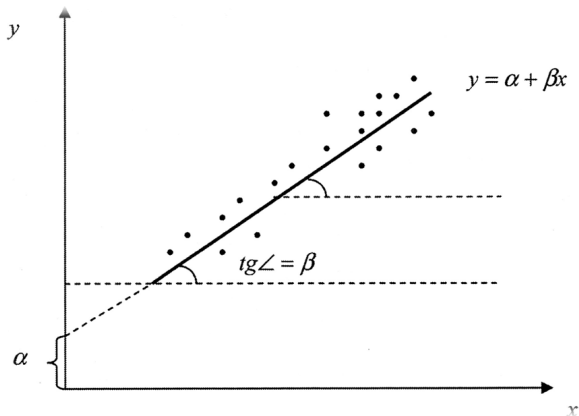
*Случайная величина e , характеризующая отклонения реального значения результата от теоретического, найденного по уравнению регрессии, называется **остатком (ошибкой)**.*

Практически парная линейная регрессия строится по данным выборки и записывается в виде:

$$y_k = a + bx_k + e_k, \quad k = 1, 2, \dots, n,$$

где n — объем выборки, $(x_k; y_k)$ — элементы выборки, a и b — выборочные оценки коэффициентов α и β регрессии, e_k — остатки.

Задача состоит в том, чтобы по данным выборки оценить коэффициенты парной линейной регрессии: α — пересечение и β — наклон линии регрессии.



Параметры α и β должны быть оценены так, чтобы линия $y = \alpha + \beta x$ наилучшим образом подходила к данным. Пусть наилучшая функция имеет вид:

$$\hat{y} = a + bx.$$

тогда качество подобранной линии можно измерить следующим образом:

$$\sum_{k=1}^n (y_k - \hat{y}_k)^2 = \sum_{k=1}^n e_k^2.$$

Это приводит к оценке параметров регрессии **методом наименьших квадратов**.

Обозначим сумму остатков как RSS (residual sum of squares):

$$RSS = \sum_{k=1}^n e_k^2 = \sum_{k=1}^n (y_k - a - bx_k)^2 \rightarrow \min.$$

Это выражение минимизируется при равенстве нулю частных производных:

$$\frac{\partial (\sum_{k=1}^n e_k^2)}{\partial a} = -2 \sum_{k=1}^n e_k = 0,$$

$$\frac{\partial (\sum_{k=1}^n e_k^2)}{\partial b} = -2 \sum_{k=1}^n e_k x_k = 0.$$

Отсюда,

$$\sum_{k=1}^n y_k - na - b \sum_{k=1}^n x_k = 0,$$

$$\sum_{k=1}^n x_k y_k - a \sum_{k=1}^n x_k - b \sum_{k=1}^n x_k^2 = 0.$$

Получили систему нормальных уравнений:

$$\begin{cases} an + b \sum_{k=1}^n x_k = \sum_{k=1}^n y_k, \\ a \sum_{k=1}^n x_k + b \sum_{k=1}^n x_k^2 = \sum_{k=1}^n x_k y_k. \end{cases}$$

Решая которую относительно a и b , получаем:

$$a = \frac{\sum_{k=1}^n y_k - b \sum_{k=1}^n x_k}{n} = \bar{y} - b\bar{x},$$

$$b = \frac{n \sum_{k=1}^n x_k y_k - \sum_{k=1}^n x_k \sum_{k=1}^n y_k}{n \sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k \right)^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}.$$

Найденная критическая точка является точкой минимума функции RSS . В этом легко убедиться, вычислив определитель из вторых производных и убедившись в его неотрицательности.

Определение

Параметр b называется **коэффициентом регрессии**. Его величина показывает среднее изменение результата y с изменением фактора x на единицу.

Через величину b выражается **средний коэффициент эластичности**

$$\Xi = b \frac{\bar{x}}{\bar{y}},$$

показывающий на сколько процентов в среднем изменится результат y при увеличении фактора x на 1%.

При положительных величинах y и x знак **свободного члена регрессии** a имеет следующую интерпретацию:

Если $a > 0$, то относительное изменение результата происходит медленнее чем относительное изменение фактора. И наоборот, если $a < 0$, то относительное изменение результата происходит быстрее чем относительное изменение фактора.

Сравним относительные изменения:

$$\begin{aligned}\frac{\Delta x}{x} \vee \frac{\Delta y}{y} &\Leftrightarrow \frac{y}{x} \vee \frac{\Delta y}{\Delta x} \Leftrightarrow \frac{a + bx}{x} \vee \frac{b\Delta x}{\Delta x} \Leftrightarrow \\ &\Leftrightarrow a + bx \vee bx \Leftrightarrow a \vee 0.\end{aligned}$$

Существует и другая интерпретация параметра a . Если $a > 0$, то вариация результата y не превосходит вариации фактора x . И наоборот, если $a < 0$, то вариация результата y превосходит вариацию фактора x .

Напомним, что вариацией случайной величины X называется

$$VX = \sqrt{DX}/EX.$$

Для определения тесноты линейной связи в модели парной линейной регрессии вычисляют **выборочный коэффициент корреляции**

$$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{S_y^2 \cdot S_x^2}}.$$

Между коэффициентом корреляции и коэффициентом регрессии существует простая взаимосвязь:

$$r_{yx} = b \cdot \sqrt{\frac{S_x^2}{S_y^2}}.$$

Если провести предварительную стандартизацию переменных, входящих в модель, т. е. выполнить нормировку

$$y^* = \frac{y - \bar{y}}{\sqrt{S_y^2}}, \quad x^* = \frac{x - \bar{x}}{\sqrt{S_x^2}},$$

то уравнение парной линейной регрессии имеет вид

$$\hat{y}^* = r_{yx} \cdot x^*.$$

Это так называемое **уравнение в стандартизированной форме**.

Оценку качества уравнения парной линейной регрессии проводят при помощи **коэффициента детерминации** r_{yx}^2 :

$$R^2 = r_{yx}^2 = \frac{(\overline{xy} - \bar{x} \cdot \bar{y})^2}{S_y^2 \cdot S_x^2} = b^2 \frac{S_x^2}{S_y^2}.$$

Для оценки качества уравнения парной линейной регрессии также используют **среднюю ошибку аппроксимации**

$$\bar{A} = \frac{1}{n} \sum_{k=1}^n \left| \frac{y_k - \hat{y}_k}{y_k} \right| \cdot 100\%.$$

Допустимое значение \bar{A} — 8–10%.

СПАСИБО ЗА ВНИМАНИЕ!