



Нелинейные модели парной регрессии

Реальные экономические процессы наилучшим образом описываются нелинейными соотношениями.

Нелинейными зависимостями описываются многие экономические законы – *спроса и предложения, потребления, производственные функции* и др.

Различают два класса нелинейных моделей:

1. *регрессии, нелинейные относительно объясняющих переменных, но линейные по оцениваемым параметрам.*
2. *регрессии, нелинейные по оцениваемым параметрам.*

Нелинейные модели путем соответствующих преобразований сводятся к линейным, т.е. выполняется их *линеализация*.

Регрессии, нелинейные относительно объясняющих переменных, например:

– *полиномы различных степеней:*

$$\hat{y}_x = b_0 + b_1 \cdot x + b_2 \cdot x^2,$$

$$\hat{y}_x = b_0 + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3;$$

– *равносторонняя гипербола:*

$$\hat{y}_x = b_0 + b_1/x;$$

– *полулогарифмическая функция:*

$$\hat{y}_x = b_0 + b_1 \cdot \ln x.$$

Такие модели приводятся к линейному виду простой *заменой переменных*, а дальнейшая оценка параметров производится с помощью МНК.

1. Парабола $\hat{y}_x = b_0 + b_1 \cdot x + b_2 \cdot x^2$ приводится к линейному виду с помощью замены:

$$x = x_1 \quad x^2 = x_2.$$

В результате приходим к двухфакторному уравнению

$$\hat{y}_x = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2,$$

оценка параметров которого при помощи МНК приводит к системе нормальных уравнений:

$$\begin{cases} b_0 \cdot n + b_1 \sum x_1 + b_2 \sum x_2 = \sum y \\ b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 \cdot x_2 = \sum x_1 \cdot y \\ b_0 \sum x_2 + b_1 \sum x_1 \cdot x_2 + b_2 \sum x_2^2 = \sum x_2 \cdot y \end{cases}$$

или в матричном виде

$$Y = (Z^T Z)^{-1} Z^T Y$$

В данной спецификации параметры имеют следующий экономический смысл:

b_0 – значение зависимой переменной y при значении регрессора $x=0$;

b_1 – прирост зависимой переменной y при увеличении значения регрессора x на единицу роста (скорость роста);

b_2 – скорость изменения скорости (ускорение роста);

b_3 – изменение ускорения и т.д.

2. Гиперболическая регрессия (обратная модель) вида

$$\hat{y}_x = b_0 + b_1/x$$

приводится к линейному уравнению простой заменой: $z = 1/x$. Система линейных уравнений при применении МНК будет выглядеть следующим образом:

$$\begin{cases} nb_0 + b_1 \sum \frac{1}{x_i} = \sum y_i; \\ b_0 \sum \frac{1}{x_i} + b_1 \sum \frac{1}{x_i^2} = \sum \frac{1}{x_i} y_i. \end{cases} \text{ или } \begin{cases} nb_0 + b_1 \sum z_i = \sum y_i; \\ b_0 \sum z_i + b_1 \sum z_i^2 = \sum z_i y_i. \end{cases}$$

Гиперболическая регрессия вида

$$\hat{y}_x = \frac{1}{b_0 + b_1 x}$$

сводится к линейной преобразованием $z = \frac{1}{\hat{y}_x}$.

Регрессии, нелинейные по параметрам, например:

– степенная – $\hat{y}_x = b_0 \cdot x^{b_1}$;

– показательная – $\hat{y}_x = b_0 \cdot b_1^x$;

– экспоненциальная – $\hat{y}_x = e^{b_0 + b_1 x}$.

– логистическая – $\hat{y}_x = \frac{b_0}{1 + b_1 \cdot e^{-b_2 x}}$,

Такие модели приводятся к линейному виду с помощью соответствующих преобразований, например, логарифмированием, а дальнейшая оценка параметров производится с помощью МНК.

Среди нелинейных моделей наиболее часто используется степенная функция $\hat{y} = b_0 \cdot x^{b_1}$, которая приводится к линейному виду логарифмированием:

$$\ln \hat{y} = \ln(b_0 \cdot x^{b_1})$$

$$\ln \hat{y} = \ln b_0 + b_1 \ln x$$

$$Y = B_0 + b_1 X,$$

где $Y = \ln \hat{y}$, $B_0 = \ln b_0$, $X = \ln x$,

т.е. МНК мы применяем для преобразованных данных:

$$\begin{cases} nB_0 + b_1 \sum X = \sum Y; \\ B_0 \sum X + b_1 \sum X^2 = \sum XY. \end{cases}$$

а затем потенцированием находим искомое уравнение.

Параметры модели $\hat{y} = b_0 \cdot x^{b_1}$ имеют следующее экономическое истолкование:

где b_0 – значение зависимой переменной y , полученное при единичном значении регрессора x .

b_1 – коэффициент эластичности \mathcal{E} переменной y по переменной x .

Коэффициент эластичности показывает, на сколько процентов измениться в среднем результат, если фактор изменится на 1%.

Формула для расчета коэффициента эластичности имеет вид:

$$\mathcal{E} = f'(x) \cdot \frac{x}{y}$$

Так как коэффициент эластичности для остальных функций не является *постоянной величиной*, а зависит от соответствующего значения фактора x , то обычно рассчитывается средний коэффициент эластичности:

$$\bar{\varepsilon} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}}.$$

Формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии:

Вид функции, y	Первая производная, y'	Средний коэффициент эластичности, $\bar{\varepsilon}$
$y = \beta_0 + \beta_1 \cdot x + \varepsilon$	β_1	$\frac{\beta_1 \cdot \bar{x}}{\beta_0 + \beta_1 \bar{x}}$
$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \varepsilon$	$\beta_1 + 2\beta_2 \cdot x$	$\frac{(\beta_1 + 2\beta_2 \cdot \bar{x}) \cdot \bar{x}}{\beta_0 + \beta_1 \bar{x} + \beta_2 \cdot \bar{x}^2}$
$y = \beta_0 + \frac{\beta_1}{x} + \varepsilon$	$-\frac{\beta_1}{x^2}$	$-\frac{\beta_1}{\beta_0 \bar{x} + \beta_1}$

Вид функции, y	Первая производная, y'	Средний коэффициент эластичности, $\bar{\varepsilon}$
$y = \beta_0 \cdot x^{\beta_1} \cdot \varepsilon$	$\beta_0 \cdot \beta_1 \cdot x^{\beta_1-1}$	β_1
$y = \beta_0 \cdot \beta_1^x \cdot \varepsilon$	$\beta_0 \cdot \ln \beta_1 \cdot \beta_1^x$	$\bar{x} \cdot \ln \beta_1$
$y = \beta_0 + \beta_1 \cdot \ln x + \varepsilon$	$\frac{\beta_1}{x}$	$\frac{\beta_1}{\beta_0 + \beta_1 \ln \bar{x}}$
$y = \frac{\beta_0}{1 + \beta_1 \cdot e^{-\beta_2 x + \varepsilon}}$	$\frac{\beta_0 \cdot \beta_1 \cdot \beta_2 \cdot e^{-\beta_2 x}}{(1 + \beta_1 \cdot e^{-\beta_2 x})^2}$	$\frac{\beta_1 \cdot \beta_2 \cdot \bar{x}}{\beta_1 + e^{\beta_2 \bar{x}}}$
$y = \frac{1}{\beta_0 + \beta_1 \cdot x + \varepsilon}$	$-\frac{\beta_0}{(\beta_0 + \beta_1 \cdot x)^2}$	$-\frac{\beta_1 \cdot \bar{x}}{\beta_0 + \beta_1 \bar{x}}$

Замечание. Возможны случаи, когда расчет коэффициента эластичности не имеет смысла. Это происходит тогда, когда для рассматриваемых признаков бессмысленно определение изменения в процентах.



Проверка адекватности нелинейной модели

Уравнение нелинейной регрессии дополняется показателем тесноты связи. В данном случае *это индекс корреляции*:

$$\rho = \sqrt{1 - \frac{Q_{ост}^2}{Q_{общ}^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}, \quad 0 \leq \rho \leq 1.$$

Чем ближе значение индекса корреляции к единице, тем теснее связь рассматриваемых признаков, тем более надежно уравнение регрессии.

Квадрат индекса корреляции называется *индексом детерминации* и характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака:

$$\rho^2 = 1 - \frac{Q_{ост}^2}{Q_{общ}^2} = \frac{Q_{факт}^2}{Q_{общ}^2}$$

Индекс детерминации ρ^2 сравнивают с коэффициентом детерминации R^2 для обоснования возможности применения линейной функции.

Чем больше кривизна линии регрессии, тем величина R^2 меньше ρ^2 .

Близость этих показателей указывает на то, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию.

Для проверки значимости уравнения регрессии в целом строят статистику F :

$$F = \frac{\rho^2}{1 - \rho^2} \cdot \frac{n - k - 1}{k} \sim F(\alpha; \nu_1 = k; \nu_2 = n - k - 1), \text{ где}$$

n – число наблюдений, k – число параметров при переменной x .

Пример. По данным проведенного опроса восьми групп семей известны данные связи расходов населения на продукты питания с уровнем доходов семьи.

Расходы на продукты питания, Y , тыс. руб.	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Доходы семьи, X , тыс. руб.	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7

Предположим, что связь между признаками носит нелинейный характер, и найдем параметры следующих нелинейных уравнений:

$$y = b_0 + b_1 \ln x + \varepsilon,$$

$$y = b_0 + b_1 \sqrt{x} + \varepsilon \text{ и}$$

$$y = b_0 \cdot x^{b_1} \cdot \varepsilon.$$

1. Для нахождения параметров регрессии $\hat{y}_x = b_0 + b_1 \ln x$ делаем замену $z = \ln x$. Получаем линейное уравнение регрессии:

$$\hat{y}_x = 0,305 + 1,063 \cdot \ln x.$$

2. Для нахождения параметров регрессии $\hat{y}_x = b_0 + b_1 \sqrt{x}$ делаем замену $z = \sqrt{x}$. Получаем линейное уравнение регрессии:

$$\hat{y}_x = -0,286 + 0,931 \cdot \sqrt{x}.$$

3. Для нахождения параметров регрессии $\hat{y}_x = b_0 \cdot x^{b_1}$ линеаризацию проводят с помощью логарифмирования: $\hat{Y} = B_0 + b_1 X$, где $Y = \ln y$, $X = \ln x$, $B_0 = \ln b_0$.

Получаем уравнение регрессии: $\hat{Y} = -0,305 + 0,551X$. После потенцирования находим искомое уравнение регрессии:

$$\hat{y}_x = 0,737 \cdot x^{0,551}.$$

Сравним построенные модели:

Модель	Индекс детерминации, $R^2 (r_{xy}^2, \rho_{xy}^2)$	Средняя ошибка аппроксимации, \bar{A} , %
Линейная модель, $y_x = b_0 + b_1 x$	0,987	6,52
Полулогарифмическая модель, $y_x = b_0 + b_1 \ln x$	0,918	14,51
Модель с квадратным корнем, $y_x = b_0 + b_1 \sqrt{x}$	0,991	4,98
Степенная модель, $y_x = b_0 \cdot x^{b_1}$	0,967	4,39

Модель с квадратным корнем лучше аппроксимирует исходные данные. Но так как индексы детерминации линейной модели и модели с квадратным корнем отличаются всего на 0,004, то можно обойтись более простой линейной функцией.