

УДК 519.862.6

Д.В.Демченко, Н.В.Рутковский

ОЦЕНКИ ПАРАМЕТРОВ ГИПЕРБОЛИЧЕСКОЙ РЕГРЕССИИ

Институт электронных и информационных систем НовГУ

A new method of estimating of hyperbolic regression is suggested in this article. The formulae of estimations through probabilistic characteristics of selection are derived. The investigated numerical example shows advantages of the proposed model in comparison with the well-known models of regression.

Среди экономических моделей нелинейной регрессии наиболее простой и часто используемой является модель гиперболической регрессии [1,2], которую можно представить в следующем виде:

$$y = \alpha + \frac{\beta}{x - \gamma} + \varepsilon, \quad (1)$$

где x — объясняющая переменная, y — объясняемая переменная, ε — случайная составляющая с нулевым математическим ожиданием. Коэффициенты α , β , γ называются параметрами регрессии.

Таким образом, функция регрессии y по x

$$M(y/x) = \alpha + \frac{\beta}{x - \gamma}, \quad \text{поскольку } M(\varepsilon) = 0. \quad \text{Среднее}$$

квадратичное отклонение случайной величины ε , как правило, зависит от значения объясняющей переменной x , и мы будем предполагать его пропорциональным $\frac{1}{|x - \gamma|}$. Сформулируем основную задачу: по

данным выборки $\begin{matrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{matrix}$ найти оценки параметров

регрессии, являющиеся по возможности простыми и качественными. Непосредственное применение метода наименьших квадратов к модели (1) приводит к громоздким нелинейным уравнениям, связывающим параметры регрессии. В связи с этим желательно по-

лучить более простым путем оценки параметров гиперболической регрессии. Основной целью работы и является получение таких оценок.

Запишем модель гиперболической регрессии в мультипликативном виде:

$$y = \alpha + \frac{u}{x - \gamma}, \quad (2)$$

где u — случайный коэффициент. Приравнявая правые части уравнений (1) и (2), получим $\frac{u}{x - \gamma} = \frac{\beta}{x - \gamma} + \varepsilon$,

откуда $u = \beta + \varepsilon(x - \gamma)$. Следовательно, $M(u) = \beta$ и $D(u) = (x - \gamma)^2 \cdot D(\varepsilon)$. Учитывая предположение по поводу среднего квадратичного отклонения ε , приходим к выводу, что среднее квадратичное отклонение случайной величины u не зависит от x .

Для модели (2) требуется найти оценки параметров α, γ и математического ожидания $M(u) = \beta$

на основе значений выборки $\begin{matrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{matrix}$.

Применительно к значениям выборки модель (2) можно записать следующим образом:

$$u_i = (x_i - \gamma) \cdot (y_i - \alpha), \quad i = 1, 2, \dots, n.$$

Обозначая оценки параметров α, β, γ через a, b, c , мы можем выразить оценку дисперсии случайной величины u :

$$\frac{1}{n} \sum_{i=1}^n (u_i - b)^2 = \frac{1}{n} \sum_{i=1}^n ((x_i - c)(y_i - a) - b)^2.$$

В соответствии с этим выражением оценки a, b, c определим как МНК-оценки для функции

$$F(a, b, c) = \sum_{i=1}^n ((x_i - c)(y_i - a) - b)^2.$$

Запишем необходимые условия минимума

$$\begin{cases} \frac{\partial F}{\partial a} = -2 \sum_{i=1}^n (x_i - c)((x_i - c)(y_i - a) - b) = 0, \\ \frac{\partial F}{\partial b} = -2 \sum_{i=1}^n ((x_i - c)(y_i - a) - b) = 0, \\ \frac{\partial F}{\partial c} = -2 \sum_{i=1}^n (y_i - a)((x_i - c)(y_i - a) - b) = 0. \end{cases} \quad (3)$$

Из второго уравнения этой системы получаем равенство $\sum_{i=1}^n (x_i - c)(y_i - a) = nb$. Используя обозначение

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, запишем выражение b через a и c :

$$b = \overline{xy} - a\bar{x} - c\bar{y} + ac. \quad (4)$$

Тогда $(x_i - c)(y_i - a) - b = (x_i \cdot y_i - \overline{xy}) - a(x_i - \bar{x}) - c(y_i - \bar{y})$.

Подставляя полученное выражение в первое и третье уравнения системы (3) и учитывая равенство

$\sum_{i=1}^n ((x_i - c)(y_i - a) - b) = 0$, запишем систему двух уравнений с неизвестными a и c :

$$\begin{cases} \sum_{i=1}^n x_i ((x_i \cdot y_i - \overline{xy}) - a(x_i - \bar{x}) - c(y_i - \bar{y})) = 0, \\ \sum_{i=1}^n y_i ((x_i \cdot y_i - \overline{xy}) - a(x_i - \bar{x}) - c(y_i - \bar{y})) = 0. \end{cases}$$

Разделив оба уравнения на n и выделив неизвестные, перепишем систему в виде

$$\begin{cases} a(\overline{x^2} - \bar{x}^2) + c(\overline{xy} - \bar{x} \cdot \bar{y}) = \overline{x^2 y} - \bar{x} \cdot \overline{xy}, \\ a(\overline{xy} - \bar{x} \cdot \bar{y}) + c(\overline{y^2} - \bar{y}^2) = \overline{xy^2} - \overline{xy} \cdot \bar{y}. \end{cases}$$

Далее воспользуемся известными обозначениями для оценок дисперсии и коэффициента корреляции: $\overline{x^2} - \bar{x}^2 = \sigma_x^2$, $\overline{y^2} - \bar{y}^2 = \sigma_y^2$, $\overline{xy} - \bar{x} \cdot \bar{y} = \text{cov}(x, y) = r\sigma_x\sigma_y$, $\overline{x^2 y} - \bar{x} \cdot \overline{xy} = \text{cov}(x, xy)$, $\overline{xy^2} - \overline{xy} \cdot \bar{y} = \text{cov}(xy, y)$. Система примет вид

$$\begin{cases} a\sigma_x^2 + cr\sigma_x\sigma_y = \text{cov}(x, xy), \\ ar\sigma_x\sigma_y + c\sigma_y^2 = \text{cov}(xy, y). \end{cases}$$

Это линейная система, ее определитель $\nabla = \sigma_x^2 \cdot \sigma_y^2 - r^2 \sigma_x^2 \sigma_y^2 = \sigma_x^2 \cdot \sigma_y^2 \cdot (1 - r^2)$. Естественно считаем, что $\sigma_x \neq 0$, $\sigma_y \neq 0$ и что случайная величина y не зависит линейно от x . Последнее влечет неравенство $r^2 \neq 1$, и значит, $\nabla \neq 0$.

Таким образом, приходим к однозначным оценкам a и c , определяемых формулами

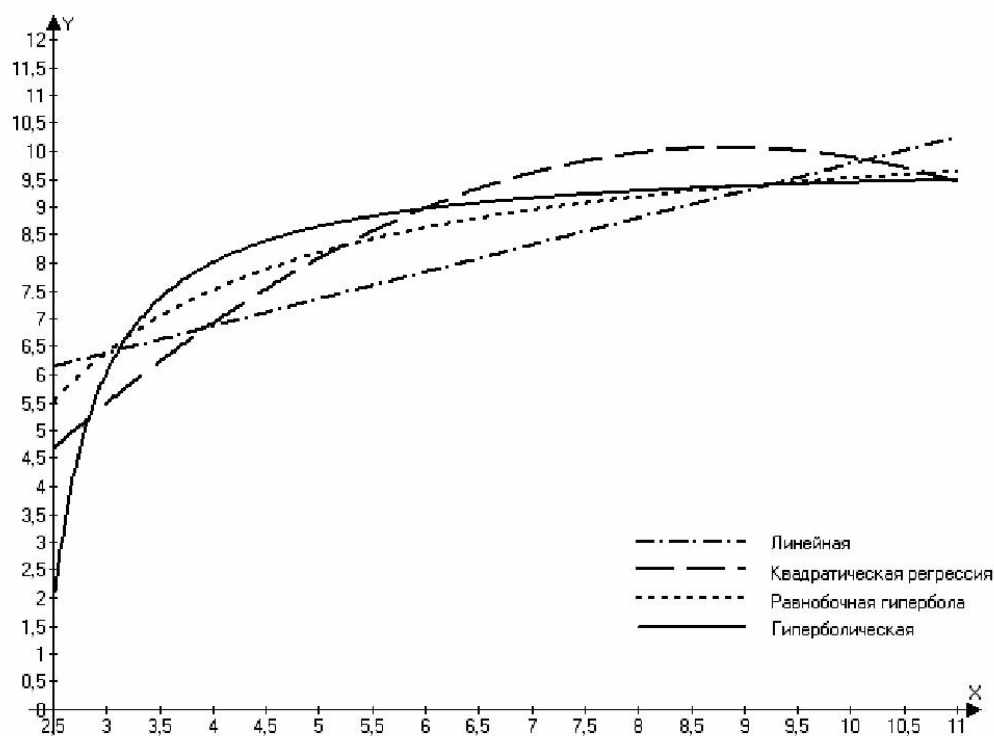
$$\begin{aligned} a &= \frac{\sigma_y^2 \text{cov}(x, xy) - r\sigma_x\sigma_y \text{cov}(xy, y)}{\sigma_x^2 \cdot \sigma_y^2 (1 - r^2)}, \\ c &= \frac{\sigma_x^2 \text{cov}(xy, y) - r\sigma_x\sigma_y \text{cov}(x, xy)}{\sigma_x^2 \cdot \sigma_y^2 (1 - r^2)}. \end{aligned} \quad (5)$$

Оценка b параметра β выражается через a и c по формуле (4).

Оставляя на будущее исследование вопроса о качестве полученных оценок, рассмотрим численный пример, на котором сравним модель (2) с оценками (4) и (5) с известными моделями линейной и квадратичной регрессий и равнобочной гиперболы (см.

Исходные данные

x	6	7	12	9	2,5	4	4,5	8	11	3	6,5	10	3,5	5
y	9	9	9,5	9,4	2	8	8,5	9,3	9,5	6	9,1	9,5	7,3	8,7



Графики уравнений регрессии

рис.). В качестве критерия сравнения моделей возьмем величину суммы квадратов отклонений (СКО) и коэффициента детерминации R^2 .

Рассматриваемые модели

- линейная регрессия: $y = a + b \cdot x + \varepsilon$;
- квадратическая регрессия: $y = a + b \cdot x + c \cdot x^2 + \varepsilon$;
- равнобочная гипербола: $y = a + b/x + \varepsilon$;
- гиперболическая регрессия: $y = a + b/(x + c) + \varepsilon$.

Полученные расчетные уравнения

- линейная регрессия:
 $\hat{y} = 4,93 + 0,48 \cdot x$ (СКО = 26,27, $R^2 = 0,51$);
- квадратическая регрессия:
 $\hat{y} = -0,45 + 2,38 \cdot x - 0,13 \cdot x^2$ (СКО = 13,05, $R^2 = 0,71$);
- равнобочная гипербола:

- $\hat{y} = 10,85 - 13,40/x$ (СКО = 13,67, $R^2 = 0,75$);
- гиперболическая регрессия:
 $\hat{y} = 9,91 - 3,71/(x - 2,04)$ (СКО = 0,075, $R^2 = 0,99$).

Вывод

Как показывают расчеты, из четырех рассмотренных моделей наиболее соответствует исходным данным гиперболическая модель. Об этом свидетельствуют самое высокое значение коэффициента детерминации и наименьшее значение суммы квадратов отклонений.

1. Колемаев В.А. Эконометрика. М.: ИНФРА-М, 2004. 150 с.
2. Charemza W.W. and Deadman D.F. New directions in econometric practice. Cheltenham, UK, 1997. 344 p.