ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 19 МАТЕМАТИКИ Выпуск 3
2012

Т. А. Макарова, **А. Н. Тырсин** (Челябинск, ЧелГУ, Екатеринбург, НИЦ «НиР БСМ» УрО РАН). **Непараметрическое оценивание линейных структурных соотношений между случайными величинами на малых выборках.**

Рассмотрим наиболее распространенный и практически важный случай одномерного структурного соотношения [1]:

$$Y = \beta_0 + \beta_1 X, \quad U = X + \xi, \quad V = Y + \varepsilon, \tag{1}$$

где коэффициенты β_0, β_1 — постоянные величины, причем $\beta_1 \neq 0$, X, ξ, ε — случайные величины, не коррелированные друг с другом и $\mathbf{E}[\xi_i] = \mathbf{E}[\varepsilon_i] = 0$, $\mathbf{D}[\xi_i] = \sigma_\xi^2$, $\mathbf{D}[\varepsilon_i] = \sigma_\varepsilon^2$ для всех i.

Будем предполагать, что распределение случайной погрешности ξ является симметричным, а ее функция распределения F_ξ имеет ограниченный носитель:

$$\mathbf{E}[\xi^3] = 0, \text{ найдется } C_1 > 0\colon\ F_\xi(x) = 0 \text{ для всех } |x| > C_1. \tag{2}$$

Считаем, что распределение случайной величины X является несимметричным, а ее функция распределения F_X имеет ограниченный снизу носитель:

$$\mathbf{E}[X^3] \neq 0, \text{ найдется } C_2\colon\ F_X(x) = 0 \text{ для всех } x < C_2. \tag{3}$$

Из условий (2) и (3) следует, что функции распределения суммы $U = X + \xi$ также будет иметь ограниченный снизу носитель: найдется такое C_3, что $F_U(x) = 0$ для всех $x < C_3$. Будем предполагать, что из физических условий рассматриваемой задачи имеется возможность построить достаточно точную оценку границы C_3, поэтому считаем ее известной.

Задача заключается в том, чтобы по результатам измерений (u_i, v_i), $i = 1, 2, \ldots, n$, оценить коэффициенты β_0 и β_1. Рассмотрим задачу оценки коэффициента β_1 отдельно. Действительно, после нахождения оценки b_1 коэффициента β_1 оценка b_0 коэффициента β_0 определяется как $b_0 = \overline{v} - b_1\overline{u}$. Введем центрированные случайные величины $X' = X - \mathbf{E}[X]$, $U' = U - \mathbf{E}[U] = X' + \xi$, $V' = V - \mathbf{E}[V] = \beta_1 X' + \varepsilon$.

Рассмотрим две оценки неизвестного коэффициента β_1:

$$b_{11} = \arg\min_b \frac{1}{n}\sum_{i=1}^{n}(v_i' - bu_i')^2, \quad b_{12} = \arg\min_b \frac{1}{n}\sum_{i=1}^{n}(u_i' + C)(v_i' - bu_i')^2, \tag{4}$$

где $u_i' = u_i - \overline{u}$, $v_i' = v_i - \overline{v}$, $i = 1, 2, \ldots, n$, $C = C_3 - \overline{u}$, b_{11}, b_{12} — оценки наименьших и взвешенных наименьших квадратов, причем в силу выбора константы C веса $(u_i' + C)$ в (4) положительны.

Теорема. *Пусть дана модель* (1) *и выполнены условия* (2), (3). *Математические ожидания оценок* b_{11} *и* b_{12} *совпадают тогда и только тогда, когда* $\sigma_\xi^2 = 0$.

Д о к а з а т е л ь с т в о. Из модели (1) следует

$$\sigma_U^2 = \sigma_X^2 + \sigma_\xi^2. \tag{5}$$

Известно [1], что математическое ожидание оценки b_{11} равно

$$\mathbf{E}\,[b_{11}] = \beta_{11} = \beta_1 \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\xi^2}. \tag{6}$$

Найдем математическое ожидание $\mathbf{E}\,[b_{12}] = \beta_{12}$. Оценка b_{12} находится из уравнения $Q \equiv n^{-1}\sum_{i=1}^{n} u_i'(u_i' + C)(v_i' - b_{12}u_i') = 0$. Переходя к математическим ожиданиям в последнем уравнении, получим $\mathbf{E}\,Q = \beta_1(\mu_3(U) + C\sigma_X^2) - \beta_{12}(\mu_3(U) + C\sigma_U^2) = 0$, где $\mu_3(U) = \mathbf{E}\,[U'^3]$. Отсюда имеем

$$\beta_{12} = \beta_1 \frac{\mu_3(U) + C\sigma_X^2}{\mu_3(U) + C\sigma_U^2}. \tag{7}$$

Из (6) и (7) с учетом условий (2) и (3) получим

$$\beta_{12} - \beta_{11} = \beta_1 \frac{\mu_3(U)\sigma_\xi^2}{\sigma_U^2(\mu_3(U) + C\sigma_U^2)} = \beta_1 \frac{\mathbf{E}\,[X^3]\sigma_\xi^2}{\sigma_U^2(\mathbf{E}\,[X^3] + C\sigma_U^2)},$$

следовательно, $\beta_{11} = \beta_{12}$ тогда и только тогда, когда $\sigma_\xi^2 = 0$. Если $\sigma_\xi^2 = 0$, то значения случайной величины X доступны без погрешности $(U = X)$, и обе оценки b_{11} и b_{12} являются несмещенными, т. е. $\beta_{11} = \beta_{12}$. Теорема доказана.

Решив систему уравнений (5)–(7), получим

$$\beta_1 = \frac{\beta_{12}(\mu_3(U) + C\sigma_U^2) - C\beta_{11}\sigma_U^2}{\mu_3(U)}. \tag{8}$$

Оценивание коэффициентов модели (1) и дисперсий σ_X^2 , σ_ξ^2 по выборке осуществим путем подстановки в (8) вместо теоретических величин их выборочных оценок:

$$b_1 = \frac{b_{12}(\overline{u'^3} + C\overline{u'^2}) - Cb_{11}\overline{u'^2}}{\overline{u'^3}}, \quad b_0 = \overline{v} - b_1\overline{u}, \quad s_X^2 = \frac{b_{11}s_U^2}{b_1}, \quad s_\xi^2 = s_U^2 - s_X^2,$$

где $\overline{u'^2}$, $\overline{u'^3}$ – выборочные оценки второго и третьего центрального моментов, b_{11}, b_{12} — оценки, вычисленные по формулам (4).

Аналогичным образом можно решить задачу построения линейного многомерного структурного соотношения.

Предложенный метод, в отличие от решения, описанного в [2], требующего оценивать четвертый центральный момент, ограничивается определением третьего центрального момента. Это позволяет существенно повысить точность оценивания на малых выборках.

СПИСОК ЛИТЕРАТУРЫ

1. *Дрейпер Н.*, *Смит Г.* Прикладной регрессионный анализ. Кн. 1. М.: Финансы и статистика, 1986.
2. *Тимашев С. А.*, *Тырсин А. Н.* Оценивание линейных структурных соотношений между случайными величинами. — Заводская лаборатория. Диагностика материалов, 2010, т. 76, № 3, с. 68–71.