

А. А. В а р и к о в (Уфа, ИМсВЦ РАН, УГАТУ). **Определение параметров модифицированной непараметрической модели для расчета налоговых обязательств.**

В работе [1] изложены основные этапы построения модифицированной непараметрической модели для расчета налоговых обязательств. Суть модели состоит в оптимальной кластеризации исследуемой совокупности, учете силы влияния каждого фактора, непрерывном сращивании модели на стыках кластеров.

Кластеризация в модели ведется по масштабным факторам. Пусть, например, $x_1, x_2, \dots, x_e$ — масштабные факторы. Тогда масштаб субъекта налогообложения (СН) определяем по формуле $r = r(x) = (\sum_{k=1}^e a_k x_k^2)^{1/2}$, где a_k — коэффициенты линейной модели. Далее СН ранжируются по масштабу, и одномерная совокупность чисел $r_i = (\sum_{k=1}^n a_k (x_{ik})^2)^{1/2}$ по методу k -средних разбивается на заданное число кластеров. Затем строится расширение $\tilde{B}_j$ кластеров B_j ($j = i, \dots, m$) путем присоединения заданного числа, например, 10% ближайших точек из кластеров B_{j-1} и B_{j+1} .

Расширение кластеров делается для того, чтобы построить непрерывную модель за счет сращивания подмоделей на соседних кластерах. Пусть j и $j+1$ — соседние кластеры. Обозначим $\tilde{r}_{jн}$ и $\tilde{r}_{jв}$ нижнюю и верхнюю границы расширенного кластера. Тогда

$$\hat{Y}(x) = \begin{cases} \hat{Y}(x, \tilde{B}_j), & \text{если } \tilde{r}_{(j-1)в}^2 \leq r(x) \leq \tilde{r}_{(j+1)н}^2, \\ \hat{Y}(x, \tilde{B}_{j+1}), & \text{если } \tilde{r}_{jв} \leq r(x) \leq \tilde{r}_{(j+2)н}^2, \\ z_j + (z_{j+1} - z_j) \frac{r - r_{(j+1)н}}{\tilde{r}_{jв} - \tilde{r}_{(j+1)н}}, & \text{если } \tilde{r}_{(j+1)н} \leq r(x) \leq \tilde{r}_{jв}, \end{cases}$$

где $j = 2, \dots, m-1$, $z_j = \hat{Y}(\tilde{r}_{(j+1)н} x / r(x), \tilde{B}_j)$, а $z_{j+1} = \hat{Y}(\tilde{r}_{jв} x / r(x), \tilde{B}_{j+1})$.

Число кластеров в модели можно выбирать двумя способами: в первом варианте число кластеров задается пользователем (экспертом), во втором случае выбирается оптимальное число кластеров, минимизирующее функционал

$$\Phi(k) = \sum_{j=1}^k \frac{d_j^2(\hat{B}_j)}{\hat{n}_j}, \quad (1)$$

где $\hat{B}_j$ — совокупность наблюдений из расширенного кластера $\tilde{B}_j$, оставшихся в выборке, $\hat{n}_j$ — число наблюдений из $\hat{B}_j$, а $d_j(\hat{B}_j)$ — функция вида (1), рассчитанная по множеству $\hat{B}_j$ при оптимальном (минимизирующем) параметре степени сглаживания b_j .

На каждом расширенном кластере строится предварительная модель, где $\omega(X, X_i)$ определяется по формуле:

$$\omega(X, X_i) = \left(\sum_{k=1}^n \frac{a_{jk}^2}{1 + b \sum (x_k - x_{ik})^2} \right)^{1/2},$$

(здесь a_k — коэффициенты линейной модели построенной на расширенном кластере B_j ($j = i, \dots, m$)). Таким образом, построенная функция осреднения учитывает силу влияния каждого фактора x_k ($k = 1, \dots, n$).

Выбор параметра b на каждом кластере осуществляется путем минимизации функции $d^2(b)$, определяемой по формуле:

$$d^2(b) = n^{-1} \sum_{i=1}^n \left(y_i - \sum_{j \neq i} \omega(x_j, x_i) y_j / \sum_{j \neq i} \omega(x_j, x_i) \right)^2,$$

которая задает асимптотически несмещенную оценку величины среднеквадратической погрешности непараметрической аппроксимации.

СПИСОК ЛИТЕРАТУРЫ

1. *Вариков А. А.* Применение информационных технологий для расчета налоговых обязательств. — В сб.: Труды второй Всероссийской научной конференции «Математическое моделирование и краевые задачи». Ч. 2. Самара: СамГТУ, 2005, с. 60–63.
2. *Голычев И. И., Вариков А. А.* Моделирование налоговых обязательств. — Обозрение прикл. и промышл. матем., 2005, т. 12, в. 2, с. 335–336.
3. *Хардле В.* Прикладная непараметрическая регрессия. М.: Мир, 1993, 349 с.