РОБАСТНОСТЬ РЕГРЕССИОННОГО ПРОГНОЗИРОВАНИЯ ПРИ НАЛИЧИИ ФУНКЦИОНАЛЬНЫХ ИСКАЖЕНИЙ ИНТЕРВАЛЬНОГО ТИПА МОДЕЛИ МНОГОМЕРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

Маевский В.В.

Белорусский государственный университет, Минск

Параметрическое регрессионное прогнозирование широко распространено при решении задач прогнозирования при моделировании экономических процессов [1]. Данный подход "хорошо работает" лишь в ситуации, когда гипотетическая регрессионная модель наблюдений точно выполняется для реальных данных. В прикладных задачах, к сожалению, гипотетические модельные предположения обычно нарушаются [2, 3]. Доклад посвящен исследованию регрессионного прогнозирования при функциональных искажениях интервального типа гипотетической функции регрессии для многомерной модели. Случай одномерной модели исследован в работе [4].

Рассматривается следующая многомерная модель: Пусть N-мерные наблюдения $x^t \in 5^N$ удовлетворяют уравнению:

$$x^t = \sum_{i=1}^{m} \theta^0_{(i)}\, \psi_i(u_t) + \lambda(u_t) + \xi^t = \theta^0 \psi(u_t) + \lambda(u_t) + \xi^t, \tag{1}$$

где $x^t \in 5^N$, $t \in 1$, $u_t \in 5^M$ - вектор факторов, $\{\psi_i(\cdot) : 5^M \quad 5\}$ - заданный набор m линейно независимых функций, причем $\det(\sum_{t=1}^{T} (\psi(u_t)\psi(u_t) \ )) \quad 0$, $\theta^0 = (\theta^0_{ij}) = (\theta^0_{(i)}) \in 5^{N\times m}$ - $(N\times m)$-матрица неизвестных истинных значений параметров модели, $\theta^0_{(i)} \in 5^N$, $i = 1,...,m$, $\lambda(\cdot) : 5^M \quad 5^N$ - неизвестная неслучайная вектор-функция, описывающая функциональное искажение. Случайные ошибки $\{\xi^t\}$, $\xi^t \in 5^N$, предполагаются н.о.р. случайными векторами, причем $E\{\xi^t\} = 0$, $\mathrm{Cov}\{\xi^t,\xi^t\} = \Sigma$.

Определим интервальные многомерные функциональные искажения (ИМФИ) $\lambda(\cdot)$ в (1): $\varepsilon_{i-}(u) \quad \lambda_i(u) \quad \varepsilon_{i+}(u)$, $i = 1,...,N$, где $\varepsilon_{\pm}(u) = (\varepsilon_{i\pm}(u))$ - некоторые заданные граничные вектор-функции.

Рассмотрим алгоритм прогнозирования значения $x_{T+\tau} \in 5^N$ на основе T наблюдений $x^t \in 5^N$, $t = 1,...,T$ и значений факторов $u_1,..., u_T, u_{T+\tau}$, определяемый статистикой $S(\cdot)$:

$$\hat{x}^{T+\tau} = S(x^1, ..., x^T; u_1, ..., u_T, u_{T+\tau}) \in 5^N. \qquad (2)$$

Риском статистического прогноза (2) по T наблюдениям на "глубину" τ назовем суммарную среднеквадратическую ошибку прогнозирования:

$$r(T,\tau) = E\{(\hat{x}^{T+\tau} - x^{T+\tau}) \ \ (\hat{x}^{T+\tau} - x^{T+\tau})\} \quad 0. \qquad (3)$$

Гарантированным риском прогнозирования называется точная верхняя граница риска прогнозирования [5]:

$$r_+(T,\tau) = \sup_{\lambda(u_1,...,u_T,u_{T+\tau})} r(T,\tau). \qquad (4)$$

Введем обозначения: $C = \sum_{t=1}^{T} x^t\psi \ \ (u_t) \in 5^{N\times m}$, $A = \sum_{t=1}^{T} \psi(u_t)\psi \ \ (u_t) \in 5^{m\times m}$, $\psi(u) = (\psi_i(u)) \in 5^m$, $(z)_+ = \max(z, 0)$, $\alpha_t = \psi \ \ (u_t)A^{-1}\psi(u_{T+\tau})$, $t = 1, ..., T$.

На практике в параметрическом прогнозировании наибольшее распространение получил алгоритм МНК-прогнозирования [1]:

$$\hat{x}_{T+\tau} = \hat{\theta}\,\psi(u_{T+\tau}),\ \hat{\theta} = CA^{-1}, \qquad (5)$$

где $\hat{\theta} \in 5^{N\times m}$ - МНК-оценка матрицы параметров θ^0.

Теорема. *Для искажений ИМФИ гарантированный риск (4) алгоритма (5) равен:* $r_+(T,\tau) = r_0(T,\tau) + \sum_{k=1}^{N} \max\left(\sum_{t=1}^{T} ((\alpha_t)_+\varepsilon_{k\pm}(u_t) - (-\alpha_t)_+\varepsilon_{k\mp}(u_t)) - \varepsilon_{k\mp}(u_{T+\tau})\right)^2$.

Литература

1. Четыркин Е. Статистическое прогнозирование. М.: Наука, 1977.

2. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль М. Робастность в статистике. Подход на основе функций влияния. М.: Мир, 1989.

3. Хьюбер П. Робастность в статистике. М.: Мир, 1984.

4. Харин Ю. С., Маевский В. В. Робастность регрессионного прогнозирования при наличии функциональных искажений модели. // Автоматика и телемеханика, №11. Москва, 2002, с. 118 - 137.

5. Kharin Yu. S. Robustness in statistical pattern recognition. Dordrecht: Kluwer Academic Publishers, 1996.