УДК 519.2 : 62-506.2.001

В.В. МАЕВСКИЙ

РОБАСТНОСТЬ РЕГРЕССИОННОГО ПРОГНОЗИРОВАНИЯ ПРИ НАЛИЧИИ ФУНКЦИОНАЛЬНЫХ ИСКАЖЕНИЙ МОДЕЛИ МНОГОМЕРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

The regression forecasting under functional distortions of the multivariate linear regression model is analysed in this paper. The guaranteed risk and coefficient of robustness for the LS-method for two types of functional distortions are given.

Параметрическое регрессионное прогнозирование широко распространено при решении задач прогнозирования в технике, экономике, медицине, экологии и других областях [1, 2]. Такой 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 [3, 4].

Настоящая статья посвящена исследованию регрессионного прогнозирования при функциональных искажениях гипотетической функции регрессии для многомерной модели. Случай одномерной модели исследован в работе [5].

1. Многомерная регрессионная модель и ее функциональные искажения

Пусть N-мерные наблюдения $x^t \in \mathbb{R}^N$ 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 $t \in \mathbb{N} = \{1, 2, \ldots\}$ – дискретный момент времени; $u_t \in U \in \mathbb{R}^M$ – вектор факторов в момент t; U – область «экспериментирования»; $\{\psi_i(\cdot) : \mathbb{R}^M \to \mathbb{R}\}$ – заданный набор m линейно независимых функций, причем $\det(\sum_{t=1}^{T}(\psi(u_t)\psi(u_t)')) \neq 0$; $\xi^t \in \mathbb{R}^N$ – вектор случайных ошибок наблюдения в момент t; $\theta^0 = (\theta^0_{ij}) = (\theta^0_{(i)}) \in \mathbb{R}^{N\times m}$ – $(N\times m)$-матрица неизвестных истинных значений параметров модели; $\theta^0_{(i)} \in \mathbb{R}^N$, $i = 1, \ldots, m$, $\lambda(\cdot) : \mathbb{R}^M \to \mathbb{R}^N$ – неизвестная неслучайная вектор-функция, описывающая функциональное искажение; нотация $A \in \mathbb{R}^{N\times m}$ означает, что A является $(N\times m)$-матрицей с элементами из $\mathbb{R}$. Случайные ошибки $\{\xi^t\}$ предполагаются независимыми в совокупности, одинаково распределенными случайными векторами, причем

$$E\{\xi^t\} = 0, \quad \mathrm{Cov}\{\xi^t, \xi^t\} = \Sigma.$$

Отметим, что при $u_t \equiv t$ имеем векторную трендовую модель, часто применяемую в прогнозировании [1].

Задача регрессионного прогнозирования состоит в построении некоторого прогноза $\hat{x}^{T+\tau} \in \mathbb{R}^N$ для неизвестного истинного значения $x^{T+\tau} \in \mathbb{R}^N$ в будущий момент времени $t = T + \tau$ ($\tau \geq 1$ – заданный «горизонт прогнозирования») на основе T наблюдений $x^t \in \mathbb{R}^N$, $t = 1, \ldots, T$ и значения фактора $u_{T+\tau} \in U$.

Определим два типа многомерных функциональных искажений (МФИ) $\lambda(\cdot)$ в (1).

МФИ-1. Интервальные искажения: $\varepsilon_{i-}(u) \leq \lambda_i(u) \leq \varepsilon_{i+}(u)$, $u \in U$, $i = 1, \ldots, N$, где $\varepsilon_{\pm}(u) = (\varepsilon_{i\pm}(u))$ – некоторые заданные граничные вектор-функции.

МФИ-2. Относительные искажения: $|\lambda_i(u)|/|(\theta^0\psi(u))_i| \leq \varepsilon_i,\ \ i = 1, ..., N$ $u \in U$, где $\varepsilon = (\varepsilon_i) \in \mathbb{R}^N$ – уровень «относительных» искажений. Иначе говоря, относительная погрешность задания i-й компоненты вектор-функции регрессии не превосходит $\varepsilon_i \cdot 100$ %.

Частные случаи искажений МФИ-1 и МФИ-2 исследованы в работе [6].

Для каждого из указанных классов искажений будем рассматривать еще один часто встречающийся специальный частный случай «кусочных искажений», когда априорно известно, что на некотором заданном подмножестве $U_0 \subset U$ искажения отсутствуют:

$$U = U_0 \cup U_1,\ \ U_0 \cap U_1 = \varnothing,\ \ \lambda(u) = 0, u \in U_0. \qquad (2$$

2. Характеристики робастности прогнозирования

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

$$\hat{x}^{T+\tau} = S(x^1, ..., x^T; u_1, ..., u_T, u_{T+\tau}) \in \mathbb{R}^N. \qquad (3$$

Риском статистического прогноза (3) по T наблюдениям на «глубину» τ назовем суммарную среднеквадратическую ошибку прогнозирования:

$$r(T, \tau) = E\{(\hat{x}^{T+\tau} - x^{T+\tau})'(\hat{x}^{T+\tau} - x^{T+\tau})\} \geq 0. \qquad (4$$

Гарантированным риском прогнозирования называется точная верхняя граница риска прогнозирования [7]:

$$r_+(T, \tau) = \sup_{\lambda(u_1), ..., \lambda(u_T), \lambda(u_{T+\tau})} r(T, \tau), \qquad (5$$

где область значений $\lambda(\cdot)$ определяется типом функциональных искажений.

Пусть $r_0(T, \tau) > 0$ – риск статистического прогноза (3) при отсутствии функциональных искажений, определяемый (4) при $\lambda(\cdot) \equiv 0$. Будем предполагать, что в этом случае алгоритм (3) является состоятельным, так что его риск сходится справа при увеличении длительности наблюдения $T \to \infty$ $r_0(T, \tau) \to r_0 + 0$, где $r_0 = \operatorname{tr} \Sigma > 0$ – наименьшее возможное значение риска в ситуации, когда функция регрессии априорно задана и отсутствуют искажения. В связи с этим коэффициентом робастности алгоритма прогнозирования (3) аналогично [7, 8] назовем относительное приращение гарантированного риска:

$$\kappa(T, \tau) = (r_+(T, \tau) - r_0) / r_0 \geq 0. \qquad (6$$

Чем $r_+(T, \tau)$ и $\kappa(T, \tau)$ меньше, тем более устойчив алгоритм прогнозирования (3) к исследуемому типу искажений.

3. Анализ робастности МНК-прогнозирования

Введем обозначения: $C = \sum_{t=1}^{T} x^t\psi'(u_t) \in \mathbb{R}^{N\times m}$, $A = \sum_{t=1}^{T} \psi(u_t)\psi'(u_t) \in \mathbb{R}^{m\times m}$

$L = \sum_{t=1}^{T} \lambda(u_t)\psi'(u_t) \in \mathbb{R}^{N\times m}$, $\qquad G_{T+\tau} = LA^{-1}\psi(u_{T+\tau}) \in \mathbb{R}^N$, $\qquad \psi(u) = (\psi_i(u)) \in \mathbb{R}^m$

$(z)_+ = \max(z, 0)$, $I_Z(z) = \{1, z \in Z;\ 0, z \notin Z\}$, $I(z) = I_{(0, +\infty)}(z)$, $\alpha_t = \psi'(u_t)A^{-1}\psi(u_{T+\tau})$ $t = 1, ..., T$.

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

$$\hat{x}_{T+\tau} = \hat{\theta}\,\psi(u_{T+\tau}),\ \hat{\theta} = CA^{-1}, \qquad (7$$

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

В [6] с точностью до обозначений получено следующее

Утверждение 1. Риск алгоритма прогнозирования (7) для модели наблюдения (1) при наличии функциональных искажений $\lambda(\cdot)$ равен:

$$r(T,\tau) = (\mathrm{tr}\,\Sigma)(1 + \psi'(u_{T+\tau})A^{-1}\psi(u_{T+\tau})) + (G_{T+\tau} - \lambda(u_{T+\tau}))'(G_{T+\tau} - \lambda(u_{T+\tau})). \quad (8)$$

Краткая схема доказательства. Учитывая (1), (7), имеем: $\hat\theta - \theta^0 = (\sum_{t=1}^{T}(\lambda(u_t)+\xi^t)\psi'(u_t))A^{-1}$. Тогда из свойств ξ^t следует $E\{\hat\theta - \theta^0\} = LA^{-1}$, $E\{(\hat\theta-\theta^0)'(\hat\theta-\theta^0)\} = (LA^{-1})'(LA^{-1}) + (\mathrm{tr}\,\Sigma)A^{-1}$. Используя (1), (4), получаем (8).■

Из (8) видно, что при наличии искажений $\lambda(\cdot)$ в модели (1) риск является суммой трех неотрицательных составляющих: $r(T,\tau) = r_0 + r_1 + r_2$, где $r_0 = \mathrm{tr}\,\Sigma$ – риск в условиях полной априорной информации и отсутствия искажений, $r_1 = (\mathrm{tr}\,\Sigma)\psi'(u_{T+\tau})A^{-1}\psi(u_{T+\tau})$ – приращение риска, порожденное конечностью времени наблюдения T, $r_2 = (G_{T+\tau} - \lambda(u_{T+\tau}))'(G_{T+\tau} - \lambda(u_{T+\tau}))$ – приращение риска, порожденное взаимодействием двух факторов: искажениями и конечностью времени наблюдения T.

Теорема 1. Для искажений МФИ-1 гарантированный риск алгоритма (7) равен: $r_+(T,\tau) = r_0(T,\tau) + \sum_{k=1}^{N} \max\,(\sum_{t=1}^{T}((\alpha_t)_+\varepsilon_{k\pm}(u_t) - (-\alpha_t)_+\varepsilon_{k\mp}(u_t)) - \varepsilon_{k\mp}(u_{T+\tau}))^2$.

Доказательство. Из (5) в силу непрерывности (по λ) функции риска (8) и компактности множества изменения $\{\lambda(u_t)\}$ имеем: $r_+(T,\tau) = r_0(T,\tau) + Q(\lambda) = r_0(T,\tau) + \max\limits_{\varepsilon_{i-}(u_t)\le\lambda_i(u_t)\le\varepsilon_{i+}(u_t),\,u_t\in U,\,i=1,\dots,N} (G_{T+\tau} - \lambda(u_{T+\tau}))'(G_{T+\tau} - \lambda(u_{T+\tau}))$. Так как $(G_{T+\tau} - \lambda(u_{T+\tau}))'(G_{T+\tau} - \lambda(u_{T+\tau})) = \sum_{k=1}^{N}(\sum_{t=1}^{T}\lambda_k(u_t)\psi'(u_t)A^{-1}\psi(u_{T+\tau}) - \lambda_k(u_{T+\tau}))^2 = \sum_{k=1}^{N}(\sum_{t=1}^{T}\lambda_k(u_t)\alpha_t - \lambda_k(u_{T+\tau}))^2$, то максимум $Q(\lambda)$ достигается при условии максимума каждого из выражений $Q_k = (\sum_{t=1}^{T}\lambda_k(u_t)\alpha_t - \lambda_k(u_{T+\tau}))^2$, $k = 1, \dots, N$.

Максимум Q_k может быть достигнут либо на максимуме, либо на минимуме выражения $\sum_{t=1}^{T}\lambda_k(u_t)\alpha_t - \lambda_k(u_{T+\tau})$. Тогда, подбирая величины искажений $\lambda_k(u_t) \in [\varepsilon_{k-}(u_t), \varepsilon_{k+}(u_t)]$, получаем: $Q_{k1\max} = (\sum_{t=1}^{T}\alpha_t(I(\alpha_t)\varepsilon_{k+}(u_t) + I(-\alpha_t)\varepsilon_{k-}(u_t)) - \varepsilon_{k-}(u_{T+\tau}))^2$.

Аналогично для минимизации: $Q_{k2\max} = (\sum_{t=1}^{T}\alpha_t(I(\alpha_t)\varepsilon_{k-}(u_t) + I(-\alpha_t)\varepsilon_{k+}(u_t)) - \varepsilon_{k+}(u_{T+\tau}))^2$. Тогда $Q_{k\max} = \max\,(Q_{k1\max}, Q_{k2\max})$. ■

Следствие 1. Коэффициент робастности (6) имеет вид:

$$\kappa(T,\tau) = \psi'(u_{T+\tau})A^{-1}\psi(u_{T+\tau}) +$$
$$+ (\sum_{k=1}^{N}\max\,(\sum_{t=1}^{T}((\alpha_t)_+\varepsilon_{k\pm}(u_t) - (-\alpha_t)_+\varepsilon_{k\mp}(u_t)) - \varepsilon_{k\mp}(u_{T+\tau}))^2) / (\mathrm{tr}\,\Sigma).$$

Следствие 2. В случае «кусочных искажений» (2)

$$r_+(T,\tau) = r_0(T,\tau) +$$
$$+ \sum_{k=1}^{N}\max\,(\sum_{t:u_t\in U_l}((\alpha_t)_+\varepsilon_{k\pm}(u_t) - (-\alpha_t)_+\varepsilon_{k\mp}(u_t)) - \varepsilon_{k\mp}(u_{T+\tau})\,I_{U_l}(u_{T+\tau}))^2.$$

Теорема 2. Для искажений МФИ-2 гарантированный риск (5) алгоритма прогнозирования (7) равен $r_+(T,\tau) = r_0(T,\tau) + \tilde\varepsilon'\chi$, где $\tilde\varepsilon = (\varepsilon_k^2) \in \mathbb{R}^N$, $\chi = ((\sum_{t=1}^{T}|(\theta^0\psi(u_t))_k||\alpha_t| + |(\theta^0\psi(u_{T+\tau}))_k|)^2) \in \mathbb{R}^N$, $k = 1, \dots, N$.

Доказательство. Аналогично доказательству теоремы 1 имеем: $r_+(T,\tau) = \max\limits_{|\lambda_i(u_t)|/|(\theta^0\psi(u_t))_i|\le\varepsilon_i,\,u_t\in U,\,i=1,\dots,N} r(T,\tau)$. С учетом (8) в $r(T,\tau)$ от $\{\lambda(u_t)\}$ зависит лишь второе слагаемое, которое максимизируем:

$$(G_{T+\tau} - \lambda(u_{T+\tau}))'(G_{T+\tau} - \lambda(u_{T+\tau})) = \sum_{k=1}^{N}(\sum_{t=1}^{T}\lambda_k(u_t)\alpha_t - \lambda_k(u_{T+\tau}))^2 \le$$
$$\le \sum_{k=1}^{N}(\sum_{t=1}^{T}|\lambda_k(u_t)||\alpha_t| + |\lambda_k(u_{T+\tau})|)^2 \le \sum_{k=1}^{N}(\sum_{t=1}^{T}|(\theta^0\psi(u_t))_k||\alpha_t|\varepsilon_k +$$
$$+ |(\theta^0\psi(u_{T+\tau}))_k|\varepsilon_k)^2 = \sum_{k=1}^{N}(\varepsilon_k^2(\sum_{t=1}^{T}|(\theta^0\psi(u_t))_k||\alpha_t| + |(\theta^0\psi(u_{T+\tau}))_k|)^2). \ ■$$

Следствие 1. Коэффициент робастности (6) имеет вид:

$$\kappa(T, \tau) = \psi'(u_{T+\tau})A^{-1}\psi(u_{T+\tau}) + \tilde{\varepsilon}'\chi \ / \ (\mathrm{tr}\,\Sigma).$$

Следствие 2. В случае «кусочных искажений» (2) $r_+(T, \tau) = r_0(T, \tau) + \tilde{\varepsilon}'\overline{\chi}$, где $\overline{\chi} = ((\sum_{t:u_t \in U_1} \left|(\theta^0 \ \psi(u_t))_k\right| |\alpha_t| + |(\theta^0\psi(u_{T+\tau}))_k| \, I_{U_1}(u_{T+\tau}))^2) \in \mathbb{R}^N$, $k = 1, \ldots, N$.

В заключение отметим, что результаты, полученные в [3], являются частным случаем (при $N = 1$) результатов, полученных в данной статье.

1. Четыркин Е. Статистическое прогнозирование. М., 1977.
2. Перельман И.И. // Автоматика и телемеханика. М., 1994. № 3.
3. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль М. Робастность в статистике. Подход на основе функций влияния. М., 1989.
4. Хьюбер П. Робастность в статистике. М., 1984.
5. Харин Ю.С., Маевский В.В. // Автоматика и телемеханика. М., 2002. № 11. С. 118.
6. Сталевская С.Н. // Актуальные проблемы информатики: Сб. тр. 6-й Междунар. конф.: В 2 ч. Мн., 1998. Ч. 2. С. 365.
7. Kharin Yu.S. Robustness in statistical pattern recognition. Dordrecht, 1996.
8. Харин Ю.С. // Автоматика и телемеханика. М., 1982. № 11. С. 155.

Поступила в редакцию 27.06.2003.

Владислав Владимирович Маевский – аспирант кафедры математического моделирования и анализа данных. Научный руководитель – доктор физико-математических наук, профессор кафедры математического моделирования и анализа данных Ю.С. Харин.

УДК 519.872

М.А. МАТАЛЫЦКИЙ, А.В. ПАНЬКОВ

ВЕРОЯТНОСТНЫЙ АНАЛИЗ ДОХОДОВ В БАНКОВСКИХ СЕТЯХ

Investigation of the probabilistic model of the incomes change in the banking networks is presented. The closed Markov queueing network with incomes serves as model. The sets of difference-differential equations are composed for the incomes of the network systems. These sets are solved by the operating method.

Современная банковская сеть – важная сфера национального хозяйства любого государства. Действующая в нашей стране платежная система может быть кратко охарактеризована следующим образом. На верхнем уровне банковской сети находится Национальный банк (Центральный банк), ниже – коммерческие банки с их филиалами. Национальный банк регулирует денежное обращение, обладает исключительным правом эмиссии денег и осуществляет иную деятельность по регулированию кредитно-денежных отношений в государстве. Осуществление безналичных операций между субъектами хозяйствования порождает взаимные расчеты между банками, которые возникают тогда, когда плательщик и получатель обслуживаются разными банками, а также при различном кредитовании банков и перемещении наличных денег. Все межбанковские переводы средств внутри страны осуществляются единым расчетным комплексом, включающим коммуникационную инфраструктуру, средства безопасности и обработки данных, которые эксплуатируются государственным предприятием «Межбанковский расчетный центр» (ГП «МРЦ») Национального банка – оператора электронной платежной системы. По правилам, утвержденным Национальным банком, под контролем этого центра производятся электронные межбанковские расчеты по платежным транзакциям банков и их клиентов через корреспондентские счета, открывающиеся на балансе каждого банка.