ОБ АСИМПТОТИЧЕСКОМ РАЗЛОЖЕНИИ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ ЛОКАЛЬНЫХ ПРОГНОЗОВ И ЛОКАЛЬНО-МЕДИАННОГО ПРОГНОЗА

В. В. Маевский

Введение

В настоящее время актуальным является прогнозирование показателей в экономике с помощью трендовых моделей. При этом используется модель линейной по параметрам регрессии и прогнозирование осуществляется с помощью двухэтапного МНК – алгоритма [1]. Данный алгоритм хорошо работает при выполнении всех модельных предположений [2], [3], однако в приложениях в исходных данных присутствуют искажения, и точность прогноза уменьшается [2], [4]. Поэтому важно разработать алгоритмы, устойчивые к искажениям и найти их свойства и характеристики [1], [2], [5].

Локально-медианный метод прогнозирования

Рассмотрим задачу прогнозирования тренда при наличии выбросов [2], [3]. Пусть наблюдается случайная последовательность:

$$x_t = y_t + \xi_t \nu_t,\ y_t = \sum_{i=1}^{m} \theta_i^0 \psi(t) + u_t, \tag{1}$$

где $t \in \{1, 2, ... \}$, $\psi(t) = (\psi_i(t)) \in \mathbb{R}^m$ – система m линейно независимых функция, u_t – случайная ошибка наблюдения, ν_t – значение «выброса», $\xi_t \in \{0, 1\}$ – бернуллиевская случайная величина (СВ) ($\xi_t = 0$ – «выброс» отсутствует, при $\xi_t = 1$ – присутствует).

Предполагается, что $\{u_t\}$ – н.о.р. СВ, $E\{u_t\} = 0$, $D\{u_t\} = \sigma^2 < +\infty$; $\{\nu_t\}$ – н.о.р. СВ, $E\{\nu_t\} = a_t$, $D\{\nu_t\} = K\sigma^2 < +\infty$, $K \geq 0$; $\{\xi_t\}$ – н.о.р. СВ Бернулли: $P\{\xi_t = 1\} = \varepsilon$, $P\{\xi_t = 0\} = 1 - \varepsilon$, ε – значение вероятности появления «выброса» в момент времени t. Значение $\varepsilon \in [0; 0,5)$ предполагается известным. Последовательности СВ $\{u_t\}$, $\{\nu_t\}$, $\{\xi_t\}$ являются независимыми в совокупности.

Пусть n – размер подвыборки из выборки X, $m \leq n \leq T$, $S = C_T^n$ – число всех возможных подвыборок размера n из выборки $X \in \mathbb{R}^T$; наборы $\{t_1^{(s)}, t_2^{(s)}, \dots, t_n^{(s)}\} \subset \{1, 2, \dots, T\}$ – подмножества n индексов, пронумерованные по $s = 1, .., S$; $X_T^{(s)} = (x_{t1}^{(s)}, x_{t2}^{(s)}, \dots, x_{tn}^{(s)})' \in \mathbb{R}^n$ – подвыборка из выборки X_T; Ψ_T - $(T\times m)$ – матрица с элементами $\psi_j(t)$, $t = 1, \dots, T$, $j = 1, \dots, m$; $\Psi_T^{(s)} = (\varphi_j^{(s)}(t_i))$, $i = 1, \dots, n$, $j = 1, \dots, m$ – $(n\times m)$ – матрица, полученная из $(T\times m)$ – матрицы Ψ_T.

По s-й ($s = 1, \ldots, S$) подвыборке $X_T^{(s)}$ построим МНК – оценку θ^0:

$$\hat{\theta}^{(s)} = (\Psi_T^{(s)\prime}\Psi_T^{(s)})^{-1}\Psi_T^{(s)\prime}X_T^{(s)}, \qquad (2)$$

и МНК – прогноз в момент времени $T+\tau$, где τ – глубина прогноза:

$$\hat{y}_{T+\tau}^{(s)} = \hat{\theta}^{(s)\prime}\,\psi(T+\tau). \qquad (3)$$

Определим локально-медианный прогноз как выборочную медиану локальных прогнозов (3) [6]:

$$\hat{y}_{T+\tau} = \underset{1\le s\le S}{\mathrm{med}}\,\{\hat{y}_{T+\tau}^{(s)}\}. \qquad (4)$$

Асимптотическое разложение функций распределения локальных прогнозов и локально-медианного прогноза

Обозначим $g^{(s)} = \Psi_T^{(s)}(\Psi_T^{(s)\prime}\Psi_T^{(s)})^{-1}\psi(T+\tau)$, $s = 1, \ldots, S$.

В [7] получена плотность распределения (ПРВ) локального прогноза (3):

Теорема 1. В предположениях модели (1) асимптотическая ПРВ локальных прогнозов (3) при $\varepsilon\to 0$ имеет вид:

$$p_{\hat{y}_{T+\tau}^{(s)}}(z) = \varphi(z \mid y_{T+\tau};\, \sigma^2\textstyle\sum_{k=1}^{n} g^{(s)2}_{k}) + \varepsilon(\sum_{i=1}^{n}\varphi(z \mid y_{T+\tau};\sigma^2\sum_{k=1}^{n} g^{(s)2}_{k} + \sigma^2 K g^{(s)2}_{i}) -$$

$$- n\varphi(z \mid y_{T+\tau};\, \sigma^2\textstyle\sum_{k=1}^{n} g^{(s)2}_{k})) + o(\varepsilon),\ i = 1, \ldots, S, \qquad (5)$$

где $\varphi(z \mid \mu;\, \sigma^2) = (2\pi)^{-2}\sigma^{-1}\exp(-(z-\mu)^2/2\sigma^2)$ – ПРВ нормального распределения.

Следствие. В предположениях модели (1) асимптотическая функция распределения локальных прогнозов (3) при $\varepsilon\to 0$ имеет вид:

$$F_{\hat{y}_{T+\tau}^{(s)}}(z) = \Phi(z \mid y_{T+\tau};\, \sigma^2\textstyle\sum_{k=1}^{n} g^{(s)2}_{k}(T+\tau)) + \varepsilon(\sum_{i=1}^{n}\Phi(z|y_{T+\tau};\, \sigma^2\sum_{k=1}^{n} g^{(s)2}_{k}(T+\tau) +$$

$$+ \sigma^2 K g^{(s)2}_{i}(T+\tau)) - n\Phi(z \mid y_{T+\tau};\, \sigma^2\textstyle\sum_{k=1}^{n} g^{(s)2}_{k}(T+\tau))) + o(\varepsilon),\ s = 1,\ldots, S, \qquad (6)$$

где $\Phi(z \mid \mu;\, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{z}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)dx$ – гауссовская функция распределения.

Пусть локальные прогнозы (3) строятся таким образом, что являются независимы в совокупности. Например, $\hat{y}_{T+\tau}^{(1)}$ строится по $x_1, \ldots, x_n$; $\hat{y}_{T+\tau}^{(2)}$ строится по $x_{n+1}, \ldots, x_{2n}$, при этом через S обозначим число получаемых локальных прогнозов.

Теорема 2. В предположении независимости в совокупности локальных прогнозов (3), в случае нечетного S, функция распределения локально-медианного прогноза $\hat{y}_{T+\tau}$ равна:

$$F_{\hat{y}_{T+\tau}}(z) = \prod_{i=1}^{S}(1-F_{\hat{y}^{(i)}_{T+\tau}}(z)) \sum_{k=(S+1)/2}^{S} \left(\sum_{\substack{i_1,\dots,i_k \\ i_1>i_2>\dots>i_k}}^{S} \frac{F_{\hat{y}^{(i_1)}_{T+\tau}}(z)\dots F_{\hat{y}^{(i_k)}_{T+\tau}}(z)}{(1-F_{\hat{y}^{(i_1)}_{T+\tau}}(z))\dots(1-F_{\hat{y}^{(i_k)}_{T+\tau}}(z))}\right),$$

где $F_{\hat{y}^{(i)}_{T+\tau}}(z)$ – функция распределения локальных прогнозов (5), $i = 1, \dots, S$.

Доказательство. Рассмотрим j-ю порядковую статистику $y_{(j)}$, построенную по локальным прогнозам (3). Учитывая, что события $\{z > \hat{y}_{T+\tau}^{(j)}\}$ и $\{L(z) \geq j\}$, где $L(z) = \sum_{i=1}^{S} \mathbb{1}(z - \hat{y}_{T+\tau}^{(j)})$ – число локальных прогнозов, не превосходящих z, эквивалентны, то получаем, что функция распределения j-й порядковой статистики $Fy_{(j)}(z) = P\{z > y_{(j)}\} = P\{L_S(z) \geq j\} = \sum_{k=j}^{S} P\{L(z) = k\}$.

Величина $\chi_i = \mathbb{1}(z - \hat{y}_{T+\tau}^{(i)})$ является бернуллиевской случайной величиной, $p_i = P\{\chi_i = 1\} = F_{\hat{y}^{(i)}_{T+\tau}}(z)$. Величина $L(z)$ является суммой разнораспределенных бернуллиевских случайных величин. В силу независимости (по предположению) локальных прогнозов (3) из свойства характеристической функции суммы независимых случайных величин получаем:

$$f_L(t) = \prod_{i=1}^{S} f\chi_i(t) = \prod_{i=1}^{S} (1+p_i(e^{it}-1)) = (1-p_1)(1-p_2)\dots(1-p_S) +$$

$$+\sum_{k=1}^{S} (1-p_1)(1-p_2)\dots(1-p_S)e^{it}p_k / (1-p_k) +$$

$$+ \sum_{\substack{k,l=1 \\ k<l}}^{S} (1-p_1)(1-p_2)\dots(1-p_S)e^{2it}p_kp_l / (1-p_k)/(1-p_l) + \dots + p_1p_2\dots p_Se^{Sit}.$$

Тогда вероятность $P\{L(z) = k\} = \left(\sum_{\substack{i_1,\dots i_k \\ i_1>i_2>\dots>i_k}}^{S} \frac{p_{i_1}\cdots p_{i_k}}{(1-p_{i_1})\dots(1-p_{i_k})}\right) \prod_{i=1}^{S}(1-p_i)$

и функция распределения $Fy_{(j)}(z) = \prod_{i=1}^{S}(1-p_i) \sum_{k=j}^{S} \left(\sum_{\substack{i_1,\dots i_k \\ i_1>i_2>\dots>i_k}}^{S} \frac{p_{i_1}\cdots p_{i_k}}{(1-p_{i_1})\dots(1-p_{i_k})}\right)$, откуда получаем, что функция распределения j-й порядковой статистики

$$Fy_{(j)}(z) = \prod_{i=1}^{S}(1-F_{\hat{y}^{(i)}_{T+\tau}}(z)) \sum_{k=j}^{S} \left(\sum_{\substack{i_1,\dots,i_k \\ i_1>i_2>\dots>i_k}}^{S} \frac{F_{\hat{y}^{(i_1)}_{T+\tau}}(z)\dots F_{\hat{y}^{(i_k)}_{T+\tau}}(z)}{(1-F_{\hat{y}^{(i_1)}_{T+\tau}}(z))\dots(1-F_{\hat{y}^{(i_k)}_{T+\tau}}(z))}\right),$$

где $F_{\hat{y}^{(i)}_{T+\tau}}(z)$ – функция распределения локальных прогнозов (3), $i = 1, \dots, S$.

Учитывая, что в случае нечетного S медиана является $(S+1)/2$-й порядковой статистикой, получаем требуемое.

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. *Tukey J. W.* A survey of sampling from contaminated distributions // In: Contributions to Probability and Statistics, Stanford Univ. Press, 1960.
5. *Харин Ю. С.* Робастность в статистическом распознавании образов. Мн.: «Университетское», 1992.
6. *Kharin Yu., Fursa R.* Robustness of time series forecasting for distorted trend model. – In: «Computer Data Analysis and Modeling». Mn.: BSU, 1995.
7. *Маевский В. В.* Об асимптотическом разложении функции распределения локально-медианного прогноза // В сборнике: «Математические методы в финансах и эконометрика» Мн.: БГУ, 2002.

РЕШЕНИЕ ОДНОЙ ЗАДАЧИ ДВУХФАЗНОЙ ФИЛЬТРАЦИИ

А. В. Маковский

Уравнения, описывающие процессы фильтрации в пористой среде вытекают из законов фильтрации и сохранения массы фаз; в зависимости от моделируемых условий учитываются или нет упругие свойства породы, сжимаемость, неньютоновские свойства фильтрующихся фаз, гравитационные силы и т. д.

Можно выделить следующие основные модели процессов вытеснения в пористой среде: Лейбензона-Маскета, в которой предполагается полное взаимное вытеснение одной фазы другой; Баклея-Леверетта, когда неполнота вытеснения учитывается введением относительных фазовых проницаемостей; Рапопорта-Лиса, если учитываются относительные фазовые проницаемости и капиллярный скачок давления между фазами.

Рассмотрим одномерную модель Баклея-Леверетта. Примем предположения о том, что фильтрующиеся фазы несжимаемы, их вязкости постоянны, пористая среда недеформируема, гравитационными силами можно пренебречь. Для простоты будем рассматривать случай, когда начальное распределение насыщенности $s(x,t)$ однородно. Получаем следующую краевую задачу:

$$p = p(x,t); s = s(x,t); x \in [0;L], t \in [0;\infty)$$

$$\frac{\partial u}{\partial x} = 0; u = -k\left(\frac{f_1}{\mu_1} + \frac{f_2}{\mu_2}\right)\frac{\partial p}{\partial x} \tag{1}$$

$$m\frac{\partial s}{\partial t} + u\frac{\partial F}{\partial x} = 0 \tag{2}$$