ВЫБОР ТИПА ИНТЕРВЕНЦИИ НЕСТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА ПО ЗАДАННЫМ ШАБЛОНАМ

Адерейко А. Д., Жук Е. Е.
Кафедра математического моделирования и анализа данных, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: adereyko@yandex.ru, zhukee@mail.ru

Исследуемая проблема заключается в выборе шаблона, ближайшего к реализации нестационарного временного ряда, характеризуемого наличием интервенций.

Введение

В различных прикладных областях, в том числе для экономики и финансов, важно спрогнозировать не само значение цены, а ее поведение, то есть тенденцию, определяемую трендами (особенно во времена резкого повышения или понижения цены).

Производится исследование над выбором трендовой модели (шаблона) [2, 3], ближайшей к реализации $X = \{x_t\}_{t=1}^T$ длительности T нестационарного временного ряда (ВР):

$$x_t = f(t) + u_t, \tag{1}$$

где $f(\cdot)$ – реальный тренд, а случайные величины $u_{t=1}^T$ имеют смысл ошибок наблюдений и удовлетворяют условиям:

- имеют нулевые математические ожидания;
- имеют одинаковые ограниченные дисперсии;

Пусть трендовые модели $\{\Omega_l\}_{l\in S}$ задаются своими типовыми (базовыми) трендами [2, 3] $\{f_l(\cdot)\}_{l\in S}$, ($S = \{1, \ldots, L\}$ – множество номеров этих моделей, $L \geq 2$), причем

$$\sum_{t=1}^{T} f_l(t) = 0, l \in S.$$

Получим семейства сдвига:

$$f_l^h = f_l(t) + h, t = \overline{1, T}, h \in R, l \in S,$$

где h – параметр сдвига.

Для принятия решений будем использовать решающее правило (РП) по методу наименьших квадратов (МНК) [3]. В качестве меры эффективности вычислим риск (вероятность ошибочно определить ближайший к реализации шаблон).

I. Математическая модель и постановка задачи

По реализации $X = \{x_t\}_{t=1}^T$ нестационарного ВР длительности T необходимо решить, к какой трендовой модели из $\{\Omega_l\}_{l\in S}$ она «ближе». Заранее нужно определить понятие «близость», а также предложить критерий эффективности принимаемых решений [1].

По сути шаблон и является трендом из (1). Примеров шаблонов существует много, но необходимо выбрать наиболее «близкий» из них, для поиска которого предложено использовать МНК.

Примеры наиболее часто встречающихся шаблонов:

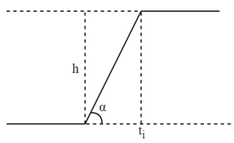

Рис. 1 – Зависит от α, h ($f_{\alpha,h}$)

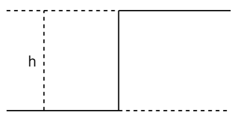

Рис. 2 – Зависит от α, h ($f_{\alpha,h}$)

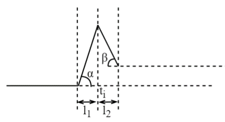

Рис. 3 – Зависит от α, β, l_1, l_2, (f_{α,β,l_1,l_2})

Для простоты вычислений можно рассмотреть только один из нескольких промежутков, содержащих интервенцию.

Пусть реализация содержит N промежутков с интервенцией, далее будет вестись работа с любым из них (например, i- м). Каждый из этих промежутков имеет точку экстремума, обозначим её t^i, тогда рассматриваемый промежуток:

$$\Delta T = [t^i - m\Delta t, t^i + n\Delta t], (m, n \geq 0)$$

разделен на $m + n$ отрезков длины Δt (всего $m + n + 1$ точка).

Пусть имеется реализация $X = \{x_t\}_{t=t^i-m\Delta t}^{t^i+n\Delta t}$ длительности $|\Delta T| = \Delta t(n+m+1)$. нестационарного временного ряда. Отсчеты $x_t \in R, t = \overline{t^i - m\Delta t, t^i + n\Delta t}, (m, n \geq 0)$ с учетом ошибок расположены возле своего тренда (см(1)).

Для (1) характерно, что

$$\{u_t\}_{t=t^i-m\Delta t}^{t^i+n\Delta t}$$

несут в себе смысл ошибок наблюдений и обладают следующими свойствами:

$$\mathrm{E}\{u_t\} = 0,$$

$$\mathrm{D}\{u_t\} = \mathrm{E}\{u_t^2\} = \sigma^2 < +\infty,$$

$$\mathrm{E}\{u_t u_l\} = 0,$$

$$\forall\, t, l = \overline{t = t^i - m\Delta t, t^i + n\Delta t},$$

$$l \neq t, m, n \geq 0$$

II. Разбор одного из шаблонов

Рассмотрим только один из шаблонов. Работа будет вестись с шаблоном, изображенным на (рис. 3). Запишем для него аналитическое представление:

$$f_{\alpha,\beta,l_1,l_2}(t) = \begin{cases} 0, \text{если } t < t^i - m\Delta t, \\ (t - t^i m\Delta t)tg\alpha, \\ \text{если } t^i - m\Delta t \leq t \leq t^i - m\Delta t + l_1, \\ -(t - t^i m\Delta t - l_1)tg\beta + l_1 tg\alpha, \\ \text{если } t^i - m\Delta t + l_1 \leq t \leq t^i + n\Delta t, \\ l_1 tg\alpha - l_2 tg\beta, \text{если } t^i + n\Delta t < t. \end{cases}$$

Данный шаблон определяет модели $\Omega_{\alpha,\beta,l_1,l_2}$.

Модель отцентрируется относительно нулевого уровня с помощью параметра сдвига h ($h \in R$):

$$f^h_{\alpha,\beta,l_1,l_2}(t) = f_{\alpha,\beta,l_1,l_2}(t) + h,$$

$$t = \overline{t^i - m\Delta t, t^i + n\Delta t}.$$

Таким образом, f_{α,β,l_1,l_2} не зависит от абсолютных значений (h) тренда.

III. Основные результаты

Обозначим $p = \{\alpha, \beta, l_1, l_2\}$ – наборы четырёх параметров, тогда

$$S = \{p : \alpha \in [-\pi, \pi], \beta \in [-\pi, \pi],$$

$$l_1 \in [0, +\infty), l_2 \in [0, +\infty)\}$$

Воспользуемся МНК для построения РП по реализации $X = \{x_t\}_{t=t^i-m\Delta t}^{t^i+n\Delta t}$ длительности $|\Delta T| = \Delta t(n+m+1)$[1, 3]. РП $d = d(X) \in S$ относит реализацию к той модели из Ω_p, к базовому тренду (шаблону) которой она ближе. Решающее правило имеет следующий вид:

$$d(X) = arg\min_p \min_{h\in R} \sum_{t=t^i-m\Delta t}^{t^i+n\Delta t} (x_t - f_p^h(t))^2 =$$

$$= arg\min_p \min_{h\in R} \sum_{t=t^i-m\Delta t}^{t^i+n\Delta t} (x_t - f_p(t) - h)^2 =$$

$$= arg\min_p \min_{h\in R} \sum_{t=t^i-m\Delta t}^{t^i+n\Delta t} (x_t - f_p(t) - \overline{x_T})^2,$$

где $\overline{x}_T = \frac{1}{\Delta T}\sum_{t=t^i-m\Delta t}^{t^i+n\Delta t} x_t$ – арифметическое (выборочное) среднее [3] значение отсчетов из реализации $X = \{x\}_{t=t^i-m\Delta t}^{t^i+n\Delta t}$, которое является решением экстремальной задачи

$$\sum_{t=t^i-m\Delta t}^{t^i+n\Delta t} (x_t - f_p - h)^2 \to \min_{h\in R}.$$

С другой стороны необходимо определить множество D° (по аналогии с [1]) номеров тех моделей из Ω_p, к которым тренд $f(\cdot)$ ближе:

$$D^\circ = \{p : \rho_*(f, f_p)\} = \min_{p\in S} \rho_*(f, f_p),$$

где

$$\rho_*^2(f, f_p) = \min_{h\in R} \sum_{t=t^i-m\Delta t}^{t^i+n\Delta t} (f(t) - \overline{f_T} - f_p)^2,$$

$\overline{f_T} = \frac{1}{\Delta T}\sum_{t=t^i-m\Delta t}^{t^i+n\Delta t} f(t)$ имеет смысл значения параметра сдвига, «приводящего» реальный тренд $f(\cdot)$ к нулевому уровню, а $\rho_*(f, f_p)$ – евклидово расстояние между трендом $f(\cdot)$ реализации X и базовым трендом $f_p(\cdot)$, определяющим модель $\Omega_p (p \in S)$.

В качестве меры эффективности используется риск [1] $r_T = P\{d(X) \notin D^\circ\}$ [1], имеющий смысл вероятности не отнести при помощи РП реализацию X к ближайшей трендовой модели. Если $D^\circ = \{d^\circ\}$, существует только одна ближайшая к тренду $f(\cdot)$ реализации X модель из $\{\Omega_p\}$ и

$$r_T = P\{d(X) \neq d^\circ\},$$

где

$$d^\circ = arg\min_p \rho_*(f, f_p)$$

Чем меньше r_T, тем эффективнее принимаемые при помощи РП решения.

Вычислим риск РП (как и в [1]) $d = d(X)$:

– истинный номер ближайшего базового тренда.

Таким образом, был найден ближайший к реализации базовый тренд для одного промежутка (i-го). После применения данного алгоритма на N промежутках сформируется оценка множества $\{d_1^\circ, d_2^\circ, \ldots, d_N^\circ\}$. И необходимый номер наиближайшего базового тренда будет оценен d^*. Поставленная задача решена.

Список литературы

1. Жук, Е. Е. Статистическое отнесение реализаций нестационарных временных рядов к заданным трендовым моделям / Е. Е. Жук // Вес. Нац. акад. навук Беларусі, Сер. фіз – мат. навук, – 2017. – № 2. – С. 52-59.
2. Андерсон, Т. Статистический анализ временных рядов: пер. с англ. / Т. Андерсон. – М: Мир, 1976. – 759 с.
3. Харин, Ю. С. Математическая и прикладная статистика / Ю. С. Харин, Е. Е. Жук. – Минск: БГУ, 2005. – 276 с.