ОПРЕДЕЛЕНИЕ ТРЕНДОВОЙ МОДЕЛИ ТИПА СДВИГА ДЛЯ НЕСТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА

Жук Е. Е., Чернявский Д. В.
Кафедра математического моделирования и анализа данных, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: zhukee@mail.com, lobiback@gmail.com

Данная работа посвящена проблеме определения ближайшей к реализации нестационарного временного ряда трендовой модели. Причем здесь трендовые модели задаются каждая своим типовым трендом, образуя семейства сдвига.

Введение

Во многих приложениях, в том числе прогнозирование поведения цены на бирже [4], важно спрогнозировать не само значение цены, а ее тенденцию, которая определяется базовыми трендами. Здесь исследуется проблема определения трендовой модели, ближайшей к наблюдаемой реализации (например, цены). Для принятия решений предлагается использовать решающее правило по методу наименьших квадратов. Для него вычислен риск, а также в качестве примера рассмотрен случай двух альтернативных трендовых моделей.

I. Математическая модель и постановка задачи

Пусть имеется реализация $X = \{x_t\}_{t=1}^T$ длительности T нестационарного временного ряда (ВР). Отсчеты $x_t \in R$, $t = \overline{1,T}$, этого ВР расположены возле своего тренда [2]:

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

где $f(t)$, $t = \overline{1,T}$, и есть тренд , а ее аргумент t интерпретируется как время. Величины $\{u_t\}_{t=1}^T$ в (1) случайны, некоррелированны, с нулевыми математическими ожиданиями и одинаковой ограниченной дисперсией:

$$\mathrm{E}\{u_t\} = 0, \quad \mathrm{D}\{u_t\} = \mathrm{E}\{u_t^2\} = \sigma^2 < +\infty; \tag{2}$$

$$\mathrm{E}\{u_t u_l\} = 0, \quad \forall\, t, l = \overline{1,T}, \quad l \neq t,$$

и имеют смысл ошибок наблюдений.

Заданы $L \geq 2$ различных трендовых моделей. Каждая такая трендовая модель Ω_l определяется своим трендом $f_l(t)$, $t = \overline{1,T}$, $l \in S$, где $S = \{1, \dots, L\}$ – множество номеров этих моделей. Все тренды $\{f_l(\cdot)\}_{l \in S}$ отцентрированы относительного нулевого уровня в том смысле, что

$$\sum_{t=1}^{T} f_l(t) = 0, \quad l \in S. \tag{3}$$

Модель Ω_l ($l \in S$) содержит разные тренды вида

$$f_l^a(t) = f_l(t) + a, \quad t = \overline{1,T}; \tag{4}$$

где $a \in R$ – параметр сдвига. Таким образом, тренд $f_l(\cdot)$ из (3) является типовым и задает «тенденцию» [4], определяющую модель Ω_l и не зависящую от абсолютных значений текущего тренда (4), принадлежащего этой модели.

Таким образом, по реализации $X = \{x_t\}_{t=1}^T$, которой соответствует ненаблюдаемый реальный тренд $f(\cdot)$ из (1), необходимо решить, к какой трендовой модели из $\{\Omega_l\}_{l \in S}$ она «ближе». Заранее нужно определить это понятие, а также предложить критерий эффективности принимаемых решений [1].

II. Решающее правило и определение его риска

Воспользуемся методом наименьших квадратов (МНК) [3] и построим по реализации $X = \{x_t\}_{t=1}^T$ решающее правило (РП) [1, 3] $d = d(X) \in S$, относящее реализацию X к той модели из $\{\Omega_l\}_{l \in S}$, к базовому тренду (3) которой она ближе. РП будет иметь вид:

$$d = d(X) = \arg\min_{l \in S} \min_{a \in R} \sum_{t=1}^{T} \left(x_t - f_l^a(t)\right)^2 = \tag{5}$$

$$= \arg\min_{l \in S} \sum_{t=1}^{T} \left(x_t - \bar{x}_T - f_l(t)\right)^2,$$

где

$$\bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t \tag{6}$$

– арифметическое среднее значений отсчетов из реализации $X = \{x_t\}_{t=1}^T$.

Определим множество $D^o \subseteq S$ номеров тех моделей из $\{\Omega_l\}_{l \in S}$, к которым реальный тренд $f(\cdot)$ ближе:

$$D^o = \left\{ k : \ \rho_*(f, f_k) = \min_{l \in S} \rho_*(f, f_l) \right\}, \tag{7}$$

где

$$\rho_*^2(f, f_l) = \min_{a \in R} \sum_{t=1}^{T} \left(f(t) - f_l^a(t)\right)^2 = \tag{8}$$

$$= \sum_{t=1}^{T} \left(f(t) - \bar{f}_T - f_l(t)\right)^2,$$

$$\bar{f}_T = \frac{1}{T} \sum_{t=1}^{T} f(t), \quad l \in S.$$

В (7), (8) $\bar{f}_T$ имеет смысл значения параметра сдвига, «приводящего» реальный тренд $f(\cdot)$ из (1) к нулевому уровню в смысле (3):

$$\sum_{t=1}^{T}(f(t)-\bar{f}_T)=0,$$

а $\rho_*(f,f_l)$ – евклидова расстояния между трендом $f(\cdot)$ реализации X (с учетом сдвига на нулевой уровень) и базовым трендом $f_l(\cdot)$, определяющим модель Ω_l ($l\in S$). В качестве меры эффективности принимаемых решений, будем использовать риск [1]

$$r_T=\mathrm{P}\{d(X)\notin D^o\}, \tag{9}$$

имеющий смысл вероятности не отнести при помощи РП $d=d(X)\in S$ из (5), (6) реализацию X к той трендовой модели, к которой она ближе в смысле базовых трендов $\{f_l(\cdot)\}_{l\in S}$ из (3).

Если $D^o=\{d^o\}$ – есть только одна ближайшая к тренду $f(\cdot)$ реализации X модель из $\{\Omega_l\}_{l\in S}$, то

$$r_T=\mathrm{P}\{d(X)\neq d^o\}. \tag{10}$$

Чем меньше (ближе к 0) значения риска r_T ($0\le r_T\le 1$) из (9), (10), тем эффективнее принимаемые при помощи РП $d=d(X)$ решения. Отметим также, что РП (5), (6), благодаря сдвигу реализации $X=\{x_t\}_{t=1}^{T}$, по которой выносится решение, на нулевой уровень:

$$\sum_{t=1}^{T}(x_t-\bar{x}_T)=0,$$

позволяет из нескольких ($m\ge 2$) реализаций $X^{(j)}=\{x_t^{(j)}\}_{t=1}^{T_j}$, $j=\overline{1,m}$, отнесенных РП (5), (6) к одной и той же трендовой модели: $d=d(X^{(1)})=\ldots=d(X^{(m)})\in S$, выбрать наиболее близкую к ней. Ее номер

$$j^*=\arg\min_{1\le j\le m}\sum_{t=1}^{T_j}\left(x_t^{(j)}-\bar{x}_{T_j}^{(j)}-f_d(t)\right)^2,$$

$$\bar{x}_{T_j}^{(j)}=\frac{1}{T_j}\sum_{t=1}^{T_j}x_t^{(j)}.$$

III. Вычисление риска

Как и в [1], вычислим риск РП $d=d(X)$ в ситуации, когда к тренду $f(\cdot)$ наблюдаемой реализации X наиболее близок лишь один из $L\ge 2$ базовых трендов $\{f_l(\cdot)\}_{l\in S}$, и

$$d^o=\arg\min_{l\in S}\rho_*(f,f_l) \tag{11}$$

– истинный номер ближайшего базового тренда через расстояния (8).

Введем обозначения:

$$\rho(f,f_l)=\sqrt{\sum_{t=1}^{T}(f(t)-f_l(t))^2},\quad l\in S;$$

$$\rho(f_l,f_k)=\sqrt{\sum_{t=1}^{T}(f_l(t)-f_k(t))^2},\quad l,k\in S,$$

– обычные евклидовы расстояния между соответствующими трендами (без учета сдвига тренда $f(\cdot)$ на нулевой уровень).

Пусть реализация $X=\{x_t\}_{t=1}^{T}$ и трендовые модели $\{\Omega_l\}_{l\in S}$ определяются соотношениями (1) – (4), а случайные величины $\{u_t\}_{t=1}^{T}$ в (2), вдобавок, независимы в совокупности и одинаково распределены по нормальному закону $N_1(0,\sigma^2)$ ($0<\sigma^2<+\infty$). Если $D^o=\{d^o\}$, где $d^o\in S$ – единственный истинный номер (11) ближайшей к X трендовой модели, то риск r_T из (10) РП (5), (6)

$$r_T= \tag{12}$$

$$=1-\mathrm{E}\left\{\prod_{\substack{l\in S\\ l\neq d^o}}U\left(z_l+\frac{\rho^2(f,f_l)-\rho^2(f,f_{d^o})}{2\sigma\,\rho(f_l,f_{d^o})}\right)\right\}$$

где $U(y)=\{1,\ y\ge 0;\ 0,\ y<0\}$ – единичная функция Хэвисайда, $\sigma=\sqrt{\sigma^2}$ – среднеквадратическое отклонение, случайные величины z_l, $l\in S$, $l\neq d^o$, имеют стандартное нормальное распределение: $\mathrm{L}\{z_l\}=N_1(0,1)$, а их совместное распределение является многомерным нормальным [3] со следующими ковариациями ($l,k\in S$, $l\neq d^o$, $k\neq d^o$):

$$\mathrm{Cov}\{z_l,z_k\}=\frac{\sum_{t=1}^{T}(f_l(t)-f_{d^o}(t))(f_k(t)-f_{d^o}(t))}{\rho(f_l,f_{d^o})\rho(f_k,f_{d^o})}.$$

Полученное соотношение (12) позволяет вычислить риск r_T РП $d=d(X)$ и оценить теоретически его эффективность. Однако, как и в [1], простой вид риск имеет лишь при $L=2$:

$$r_T=\Phi\left(-\frac{\left|\rho^2(f,f_1)-\rho^2(f,f_2)\right|}{2\sigma\rho(f_1,f_2)}\right),$$

где $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp\left(-\frac{w^2}{2}\right)dw$, $z\in R$ – функция распределения стандартного нормального закона.

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. 5259.
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.
4. Элдер, А. Как играть и выигрывать на бирже / А. Элдер. – М. : Альпина Паблишер, 2017. – 472 с.