Применение порядковых статистик в построении одношаговых прогнозов поведения финансовых индексов

В.В. Мисюра[1], Е.В. Мисюра[2]
[1]Донской государственный технический университет, Ростов-на-Дону
[2]Российский экономический университет имени Г.В. Плеханова, Москва

Существует большое количество методов и способов получения точечных и интервальных прогнозов финансовых индексов, в которых используются математические модели случайного процесса и статистические оценки параметров модели, с помощью которых модель настраивается на конкретную реализацию. Однако, отсутствие стационарности в развитии таких процессов, наличие избыточной или необъясняемой волатильности, выбросов, тяжелых хвостов, кластеризации волатильности, не позволяют найти оценки параметров модели, обладающих хорошими свойствами статистических оценок. Такие модели, в которых параметры недостаточно хорошо определены, называют моделями с неопределенными параметрами.

Работа посвящена анализу временных рядов с целью предсказания их развития во времени. Предлагается метод получения одношагового интервального прогноза временного ряда основанный на конструировании наборов предиктивных и целевых переменных с помощью робастных статистик. Основная идея предложенного в статье метода заключается в том, что левую границу одношагового интервального прогноза временного ряда $h_k, \quad k=1,...,N$ формируют k первых наименьших порядковых статистик, соответственно правую границу – порядковые статистики от $k+1$ до N, взятые с весами, для нахождения которых применяется квантильная регрессия.

Рассмотрим временной ряд $h_k, \quad k=1,...,N$ как реализацию случайного процесса $\{X_k, 0 \le k \le N\}$. Пусть некоторое четное число $1<\tau<N$ определяет ширину окна процедуры сдвига временного ряда $h_k, \quad k=1,...,N$. Преобразуем одномерный ряд h_k в два многомерных набора данных по следующему алгоритму.

1. Получаем случайные последовательности при сдвиге h_k по i на один шаг и упорядочиваем их элементы по возрастанию $(h)_i^{i+\tau-1}=\{h_i, \quad h_{i+1}\dots h_{i+\tau-1}\}, \quad i=1,...,N-\tau+1$ Преобразуем последовательности $(h)_i^{i+\tau-1}$ в вектор-столбцы $H^{(1)}, H^{(2)}..., H^{(n)} \in R^{\tau}$, где $n=N-\tau+1$.
2. Получаем матрицу

$$H=\begin{pmatrix} h_1^{(1)} & h_2^{(2)} & \cdots & h_n^{(n)} \\ h_2^{(1)} & h_3^{(2)} & \cdots & h_{n+1}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ h_\tau^{(1)} & h_{\tau+1}^{(2)} & \cdots & h_N^{(n)} \end{pmatrix},$$

где $h_l^{(j)}$ порядковая статистика с рангом l в упорядоченном наборе полученном из элементов последовательности $(h)_i^{i+\tau-1} = \{h_i, \quad h_{i+1} \dots h_{i+\tau-1}\}, \quad i = 1,...,N-\tau+1$ на j итерации.

3. Транспонируем матрицу H и наименьшие элементы каждой строки переносим в матрицу $G(N-\tau+1 \times (\tau/2+1))$, наибольшие - в матрицу $U(N-\tau+1 \times (\tau/2+1))$, первый столбец матриц представляет собой вектор-столбец, состоящий из 1:

$$G = \begin{pmatrix} 1 & h_1^{(1)} & h_2^{(1)} & \cdots & h_{\tau/2}^{(1)} \\ 1 & h_2^{(2)} & h_3^{(2)} & \cdots & h_{\tau/2+1}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & h_n^{(n)} & h_{n+1}^{(n)} & \cdots & h_{\tau/2+n-1}^{(n)} \end{pmatrix}, \ U = \begin{pmatrix} 1 & h_{\tau/2+1}^{(1)} & h_{\tau/2+2}^{(1)} & \cdots & h_{\tau}^{(1)} \\ 1 & h_{\tau/2+2}^{(2)} & h_{\tau/2+3}^{(2)} & \cdots & h_{\tau+1}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & h_{\tau/2+n}^{(n)} & h_{\tau/2+n+1}^{(n)} & \cdots & h_N^{(n)} \end{pmatrix}, \ n = N-\tau+1. \ (1)$$

Таким образом мы получаем две матрицы или два набора данных предикторных переменных, на основе которых построим квантильные регрессионные модели для вычисления правой и левой границ интервального прогноза.

В качестве целевых переменных будут выступать нижняя и верхняя эмпирические границы, полученные как линейные комбинации $\{h_i, \quad h_{i+1} \dots h_{i+\tau-1}\}, \ i = 1,...,N-\tau+1$, т.е. $\theta_-^i = \varphi(h_i, \quad h_{i+1} \dots h_{i+\tau-1}), \ \theta_+^i = \varphi(h_i, \quad h_{i+1} \dots h_{i+\tau-1}), \quad i = 1,...,N-\tau+1$.

Перейдем к вопросу определения целевых переменных (θ_-^i, θ_+^i). Целевые переменные предлагается вычислять как границы доверительного интервала порядковой статистики случайной последовательности $(h)_i^{i+\tau-1} = \{h_i, \quad h_{i+1} \dots h_{i+\tau-1}\}$. Обоснованием такого выбора является приведенная ниже теорема.

Рассмотрим общий случай. Пусть $\{H_{(1)}, \quad H_{(2)} \dots H_{(\tau)}\}$ – порядковые статистики для выборки $\{H_1, \quad H_2 \dots H_\tau\}$. Обозначим квантиль уровня p через $h_p = F^{-1}(p), 0 < p < 1$, где $F_H(h)$ - неизвестная функция распределения наблюдаемой случайной величины. Справедлива следующая теорема.

Теорема. Пусть два числа r и s такие что, $P(H_{(r)} < h_p < H_{(s)}) = 1-2\alpha$ заданная доверительная вероятность и интервал $(H_{(r)}, \ H_{(s)})$ со случайными границами $H_{(r)}$ и $H_{(s)}$ включает неизвестный квантиль $h_p = F^{-1}(p), 0 < p < 1$. Тогда вероятность $P(H_{(r)} < h_p < H_{(s)})$ не зависит от неизвестной функции распределения $F_H(h)$.

Выберем в качестве порядковой статистики для $(h)_i^{i+\tau-1} = \{h_i, \quad h_{i+1} \dots h_{i+\tau-1}\}$ медиану. Тогда в качестве интервальной оценки для $Me(h)_i^{i+\tau-1}$ можно выбрать двумерную статистику вида $(h_l, \quad h_k), \quad i \le l < k \le i+\tau-1, i = 1,...,N-\tau+1$, определяющую симметричный интервал с уровнем доверия $1-2\alpha$, полагая $k = \tau - l - 1 + i$.

Поскольку $P(h_l < Me(h)_i^{i+\tau-1} < h_k) = (1\!/\!_2)^\tau \sum_{t=l}^{\tau-l} C_\tau^t = 1-2\alpha$ значение l, используя теорему Муавра-Лапласа, можно рассчитать по формуле

$$l = \left[0,5\left\{\tau+1-\sqrt{\tau}\Psi(1-\alpha)\right\}+1\right]+i,\ i=1,...,N-\tau+1, \qquad (2)$$

где $[v]$ – целая часть числа v, $\Psi = \Phi^{-1}(u)$.

Итак, будем полагать в дальнейшем, что мы располагаем результатами регистрации значений предикторных переменных $(g_1,\ \ g_2 \ldots g_{\tau/2})$ и $(u_1,\ \ u_2 \ldots u_{\tau/2})$ представленных в вектор-столбцах матриц $G(N-\tau+1\times(\tau\!/\!_2+1))$ и $U(N-\tau+1\times(\tau\!/\!_2+1))$ (см. (1)), а так же векторами целевых переменных $\theta_- = (\theta_-^i)^T$ и $\theta_+ = (\theta_+^i)^T$.

Построим квантильные регрессий с линейной функцией зависимости, т.е. предполагаем, что условные квантили определяются соотношениями:

$$Quant_\alpha(\theta_- \mid G\ ) = a_0 + a_1 g_1 + a_2 g_2 + \ldots + a_{\tau/2} g_{\tau/2} + v_\alpha, \qquad (3)$$

$$\theta_- = \hat{a}_0 + \hat{a}_1 g_1 + \hat{a}_2 g_2 + \ldots + \hat{a}_{\tau/2} g_{\tau/2} + \hat{v}_\alpha$$

$$Quant_{1-2\alpha}(\theta_+ \mid U_N) = b_0 + b_1 u_1 + b_2 u_2 + \ldots + b_{\tau/2} u_{\tau/2} + v_{1-2\alpha}. \qquad (4)$$

$$\theta_+ = \hat{b}_0 + \hat{b}_1 u_1 + \hat{b}_2 u_2 + \ldots + \hat{b}_{\tau/2} u_{\tau/2} + \hat{v}_{1-2\alpha}$$

Приведем метод оценки параметров модели (3), который также применим без дополнительных условий к модели (4).

Оценка параметров модели (3) определяется из соотношения

$$\hat{a}_0 = \underset{a\in R^{\tau/2+1}}{\arg\min} \sum_{i:\theta_- \geq \hat{\theta}_-} \alpha|v_i| + \sum_{i:\theta_- < \hat{\theta}_-} (1-\alpha)|v_i| \qquad (5)$$

Линейная модель (5) была представлена Коенкером и Бассетом [1], как обобщение простой квантили.

Задача (5) сводится к линейному программированию. Перепишем задачу (5) в виде

$$\hat{a}_0 = \underset{a\in R^{\tau/2+1}}{\arg\min} \sum_i \alpha(v_i)_+ + (1-\alpha)(v_i)_-, \qquad (6)$$

где $(\cdot)_+$ и $(\cdot)_-$ — положительная и отрицательная части числа, соответственно. Тогда положив $r = r^+ - r^-$, где r^+, r^- – положительная и отрицательная части вектора r, получаем задачу линейного программирования

$$\begin{cases} \alpha e^T r^+ + (1-\alpha)e^T r^- \to \min \\ G^T a + r^+ - r^- = \theta_- \\ (a, r+, r-) \in R^{\tau\!/\!_2+1} \times R^{2(N-\tau+1)} \end{cases} \qquad (7)$$

где e = (1, 1, . . . , 1)^T. Очевидно, что r^+_i и r^-_i не могут иметь ненулевые значения одновременно, поэтому задачи (5) и (7) эквивалентны. На выходе мы получаем решение (a, r^+, r^-), из которого нам нужен вектор a.

Литература

[1] Koenker, R. Quantile regression / R. Koenker, J. Gilbert Bassett // Econometrics. — 1978. — Vol. 46, no. 1. — Pp. 33–50.