Инженерный вестник Дона, №4 (2018)
ivdon.ru/ru/magazine/archive/n4y2018/5338

Прогнозирование тенденций развития случайных процессов с применением порядковых статистик

В.В. Мисюра

Донской государственный технический университет, Ростов-на-Дону

Аннотация: В статье приведены результаты разработки метода прогнозирования тенденций развития случайных процессов, основанного на порядковых статистиках, а именно, медиане и статистике Ходжеса-Лемана. В статье подробно рассматривается предложенный метод: даются основные определения, формулы расчета, подробное описание алгоритма. Алгоритм реализован в виде программного модуля, который имеет практическое применение и может быть использован для решения задач прогнозирования тренда финансовых временных рядов. Приведены сравнительные результаты применения метода прогнозирования для индекса РТС в случае применения медианы и статистики Ходжеса-Лемана.
Ключевые слова: порядковая статистика, случайная величина, случайный процесс, медиана, статистика Ходжеса-Лемана, прогнозирование, тренд, средняя квадратичная ошибка, средняя абсолютная ошибка.

Рассмотрим случайные процессы описывающие поведение финансовых временных рядов. Предметом настоящего исследования являются методы прогнозирования тенденции их развития во времених использованием оценок на основе порядковых статистик. Цель исследования заключается в оценке возможности использования таких методов для прогнозирования.

Для описания эволюции величин S_t, соответствующих цене некоторого финансового инструмента в момент времени *t*, обратимся к случайному процессу $h_t = (h_t)_{1 \le t \le n}$ с дискретным временем, где

$$h_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \qquad (1).$$

Под тенденцией поведения финансового инструмента будем понимать 3 типа его поведения: падение (-1); сохранение(0); рост тренда(1).

Функция, которая позволит прогнозировать тенденцию индекса на один временной период, строится по следующей формуле:

$$Trend_i = \begin{cases} -1, \theta_+^i < 0 \\ 0, (\theta_+^i > 0) \,\&\, (\theta_-^i < 0) \\ 1, \theta_-^i > 0 \end{cases} \quad (2)$$

где θ_-^i и θ_+^i – пороговые переменные функции.

Пороговые переменные вычисляются соответственно:

$$\theta_\pm^i = \eta_{(h_k)}^i \pm \alpha \sqrt{\eta_{(\sigma_k^2)}^i}\,, \quad (3)$$

где $\eta_{(h_k)}^i$ – порядковая статистика, вычисленная по последовательности h_t за k временных периода, предшествующих уровню i; α – коэффициент, настраиваемый для каждой случайной последовательности; $\eta_{(\sigma_k^2)}^i$ – порядковая статистика, вычисленная по случайной последовательности $\sigma_k^2 = (h - \eta_{(h_k)})^2$ за k временных периода, предшествующих уровню i.

Дадим понятие порядковых статистик. Если случайные величины, входящие в выборку $X_1, X_2, ..., X_n$ расположить в порядке возрастания их значений $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$, то соответствующие этим значениям случайные величины $X_{(1)}, X_{(2)}, ... X_{(n)}$ называют порядковыми статистиками. Распределение вероятностей порядковых статистик зависит от объема выборки n, номера порядковой статистики, а также от функции распределения $F(x)$ и плотности $f(x)$ наблюдаемой случайной величины X. Статистические свойства порядковых статистик, как при конечных и малых объемах выборки, так и при больших объемах выборки (то есть в асимптотике), подробно изучены и описаны в литературе (см., например, [1]).

В качестве порядковых статистик для прогнозирования в формуле (3) могут быть использованы следующие оценки: медиана, медиана Ходжеса-

Лемана, статистика Диксона, статистика Огавы, статистика Пирсона-Тьюки, статистика Кенуя и другие (см., например, [2, 3]).

В предложенном выше алгоритме прогнозирования будем использовать две порядковые статистики – медиану и медиану Ходжеса-Лемана.

Выборочная медиана вычисляется по формуле:

$$med(X_1, X_2, ..., X_n) = \begin{cases} \frac{1}{2}\left(X_{[\frac{n}{2}]} + X_{[\frac{n}{2}+1]}\right), \text{ если } \frac{n}{2} - \text{целое} \\ X_{[\frac{n+1}{2}]}, \qquad \text{если } \frac{n+1}{2} - \text{целое} \end{cases}, \tag{4}$$

где $X_{[i]}$ – i-я порядковая статистика, равная i-му по величине значению выборочного ряда $x_1 \le x_2 \le ... \le x_n$, ранжированного по возрастанию.

Порядковая статистика Ходжеса-Лемана рассчитывается по формуле (см. [3, 4]):

$$X_{H-L} = med[(X_{[i]} + X_{[j]})/2], \quad 1 \le i \le j \le n. \tag{5}$$

В формуле (5) *med*обозначает выборочную медиану. Свойства медианы Ходжеса-Лемана X_{H-L} хорошо изучены и подробно описаны в литературе [2, 3, 5].

Для решения задачи прогнозирования тенденции развития случайных процессов во времени предлагается следующий алгоритм:

1. Определение случайной последовательности $h_t = (h_t)_{1 \le t \le n}$ по формуле (1). Определение исходных данных – коэффициента α и ширины окна k.

2. Расчет пороговых значений θ_-^i и θ_+^i по формуле (3), где $i = k+1, 2, ..., n$. Вычисление показателей тренда по формуле (2).

3. Вычисление ошибки прогноза–средняя абсолютная ошибка (MAE, MeanAbsoluteError) и средняя квадратичная ошибка (RMSE, RootMeanSquareError) по формулам (6) и (7) соответственно [5].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - f_i\right| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2} \tag{7}$$

где y_i – фактическое значение в момент времени i; f_i – прогнозное значение в момент времени i; n – размер горизонта прогнозирования.

Алгоритм реализован в программном модуле VBAв MSExcel и опробован на прогнозировании индекса РТС. В качестве исходных данных для исследования использовались значения индекса РТС по дням с 20.04.2018 по 20.11.2018 (на момент закрытия торгов) [6]. На рис. 1 показана динамика значений индекса РТС.

Рис. 1. – Исходные данные

В таблице 1 приведены сравнительные результаты применения метода прогнозирования индекса РТС в случае применения медианы и статистики Ходжеса-Лемана, приведены значения ошибок при различных входных параметрах.

Таблица № 1

Результаты ошибок прогнозирования индекса РТС в случае применения медианы и статистики Ходжеса-Лемана

k	α	$\eta^{i}_{(h_k)}$ - медиана		$\eta^{i}_{(h_k)}$ - статистика Ходжеса-Лемана	
		MAE	RMSE	MAE	RMSE
6	1,3	0,18	0,175	0,09	0,085
7	1,3	0,19	0,18	0,1	0,097
6	1,96	0,086	0,077	0,074	0,065
7	1,96	0,094	0,087	0,089	0,078

Проанализировав результаты расчетов, заметим, что применение и медианы, и статистики Ходжеса-Лемана для прогнозирования дает неплохие результаты, хотя при применении порядковой статистики Ходжеса-Лемана ошибки незначительно меньше.

В статье был предложен непараметрический метод, который не использует информацию о функциональном характере распределений. Этот метод гарантирует заданное качество решений в рамках непараметрических моделей при неизвестном распределении наблюдений. Методы, основанные на использовании известного функционального вида распределения наблюдений, зачастую являются очень чувствительными к выборочным данным [7,8]. Что касается качества прогнозов, то эти вопросы рассмотрены в [9,10] и выбранные ошибки MAE и RMSE являются достаточно надежными показателями качества прогнозов.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-01-00888 А

Литература

1. Дейвид Г. Порядковые статистики. М.: Наука. Главная редакция физико-математической литературы, 1979. 336 с.
2. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
3. Васильев А.А. Гибридные модели прогнозирования объема продаж нового товара с использованием оценок на основе порядковых статистик // Современные научные исследования и инновации, 2014, № 8. Ч. 2. URL: web.snauka.ru/issues/2014/08/37268
4. Hodjes J.L., Lehmann E.L. Estimation of location based on rank tests // Ann. Math. Statist, 1963, №4, pp. 598-611.
5. Хеттманспергер Т. Статистические выводы, основанные на рангах. М.: Финансы и статистика, 1987. 334 с.
6. Значения индексов по дням // Московская биржа. Индексы. URL: moex.com/ru/index/stat/allindexdata.aspx
7. Мисюра В.В., Мисюра И.В. Обработка и фильтрация сигналов, Современное состояние проблемы // Инженерный вестник Дона, 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2130.
8. Мисюра В.В., Мисюра И.В. Программное обеспечение моделирования и фильтрации сигналов сложной нелинейной природы // Инженерный вестник Дона, 2016, №3 URL: ivdon.ru/ru/magazine/archive/n3y2016/3749.
9. Hettmansperger T.P., McKean J.W. Robust nonparametric statistical methods. 2nd ed. USA, Florida: CRC Press, 2010. 553 p.
10. Турунцева М.Ю. Инструменты оценки качества прогнозов показателей экономической деятельности организаций // Экономические отношения, 2011, № 2 (2).URL: bgscience.ru/lib/9844/

References

1. Dejvid G. Porjadkovye statistiki [Ordinal statistics]. M.: Nauka. Glavnaja redakcija fiziko-matematicheskoj literatury, 1979. 336 p.
2. Kobzar' A.I. Prikladnaja matematicheskaja statistika. Dlja inzhenerov i nauchnyh rabotnikov [Applied mathematical statistics. For engineers and scientists]. M.: FIZMATLIT, 2006. 816 p.
3. Vasil'ev A.A. Sovremennye nauchnye issledovanija i innovacii, 2014, № 8. Ch. 2. URL: web.snauka.ru/issues/2014/08/37268
4. Hodjes J.L., Lehmann E.L. Ann. Math. Statist, 1963, №4, pp. 598-611.
5. Hettmansperger T. Statisticheskie vyvody, osnovannye na rangah [Statistical conclusions based on ranks]. M.: Finansy i statistika, 1987. 334 p.
6. Moskovskaja birzha. Indeksy. URL: moex.com/ru/index/stat/allindexdata.aspx
7. Misjura V.V., Misjura I.V. Inženernyj vestnik Dona (Rus), 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2130.
8. Misjura V.V., Misjura I.V. Inženernyj vestnik Dona (Rus), 2016, №3. URL: ivdon.ru/ru/magazine/archive/n3y2016/3749.
9. Hettmansperger T.P., McKean J.W. Robust nonparametric statistical methods. 2nd ed. USA, Florida: CRC Press, 2010. 553 p.
10. Turunceva M.Ju. Jekonomicheskie otnoshenija, 2011, № 2 (2). URL: bgscience.ru/lib/9844/