

АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ АНСАМБЛЯ МОДЕЛЕЙ

Т. В. Голод

fpm.golodtv@bsu.by;

Научный руководитель – В. И. Малюгин, доктор экономических наук, доцент

В статье рассматривается задача прогнозирования однопериодных приращений курсов финансовых активов. Для решения задачи предлагается метод моделирования и прогнозирования временных рядов на основе алгоритма декомпозиции SEEMDAN временных рядов и ансамбля эконометрических моделей ARIMA. Алгоритм SEEMDAN раскладывает исходный временной ряд на отдельные компоненты (моды), для каждой из которых строится прогноз с использованием моделей ARMA и ARIMA. Итоговый прогноз представляет собой комбинацию прогнозов на основе полученного ансамбля моделей. Приводятся результаты сравнительного анализа точности прогнозов на основе используемого метода и альтернативных типов моделей, которые говорят о том, что предлагаемый метод позволяет значительно улучшить точность прогнозов.

Ключевые слова: Декомпозиция временных рядов; алгоритмы декомпозиции SEEMDAN и EMD; модели ARMA и ARIMA; ансамбль моделей.

ВВЕДЕНИЕ

Как известно, основной принцип торговли на финансовом рынке определяется «золотым правилом инвестирования»: «покупай дешево, продавай дорого» [1]. Данный принцип создает необходимость в проведении анализа рынка, целью которого является прогнозирование будущих значений курсов торгуемых активов.

Можно выделить два подхода к моделированию рыночных данных: моделирование самих нестационарных курсов либо же переход к стационарным доходностям или однопериодным приращениям и их последующее моделирование. В рамках второго подхода так или иначе все методы предполагают использование наблюдаемой в финансовых временных рядах автокорреляции.

Альтернативный подход, заключающийся в моделировании нестационарных временных рядов курсов, не может опираться на подходы, основанные на использовании линейных закономерностей динамики нестационарных временных рядов, поскольку для нестационарных временных рядов автокорреляция стремится к единице. Это обуславливает необходимость использования иных нелинейных моделей и комплексных подходов, которые способны раскрыть сущность сложной динамики.

В данной статье предлагается и исследуется метод моделирования и прогнозирования однопериодных изменений исходного временного ряда. В основе данного метода лежит идея использования ансамбля моделей на основе декомпозиции исходного временного ряда для задачи моделирования курсов, описанная в [2]. Он использует алгоритм CEEMDAN (*Complete Ensemble Empirical Mode Decomposition with Adaptive Noise*), который осуществляет декомпозицию исходного временного ряда на осциллирующие компоненты, различных порядков частот, называемые модами. По результатам разложения высокочастотные компоненты являются очевидно стационарными, что позволяет использовать классические статистические методы для их моделирования, в частности модель ARMA. Часть низкочастотных компонентов, которые в некотором смысле описывают кривую тренда, очевидно не являются стационарными и требуют моделирования с помощью других нелинейных методов. В частности, авторы статьи [2] предлагают для их моделирования использовать нейронную сеть архитектуры LSTM. Сумма прогнозов по всем компонентам образует комбинированный прогноз ансамбля моделей для исследуемого временного ряда курса.

Можно заметить, что устройство описанного подхода концептуально безразлично к статистической природе временного ряда. С помощью данного метода могут моделироваться и нестационарные курсы, что предлагается и исследуется в работе [2], и стационарные однопериодные приращения курсов. На практике с моделированием нестационарных компонентов сопряжены основные недостатки описанного подхода. Большое количество нестационарных компонентов ухудшает качество совокупного прогноза вследствие низкой точности возможного прогноза нестационарных компонентов. В связи с этим точность прогнозирования при анализе однопериодных приращений курсов будет выше, чем точность прогнозирования при анализе самих курсов, поскольку количество нестационарных компонентов в разложении стационарных однопериодных приращений курсов меньше, чем в разложении нестационарных курсов, а их динамика и форма проще. Помимо этого, по результатам экспериментов известно, что качество прогноза нейронной сети LSTM для нестационарных компонентов незначительно хуже качества прогноза модели ARIMA. При этом использование и обучение модели ARIMA значительно быстрее использования и обучения нейронной сети LSTM. Таким образом, незначительно теряя в точности возможно достичь существенного прироста в скорости работы алгоритма, что является существенным в контексте использования в реальных приложениях.

Учитывая характер недостатков метода, предложенного в статье [2], и специфику их устранения был разработан собственный метод моделирования временных рядов однопериодных приращений курсов с использованием модели ARMA для моделирования стационарных компонентов и ARIMA для моделирования нестационарных компонентов разложения рассматриваемого временного ряда.

АЛГОРИТМ ПОСТРОЕНИЯ АНСАМБЛЯ МОДЕЛЕЙ

Разработанный метод состоит из двух этапов: декомпозиция временного ряда с помощью алгоритма CEEMDAN и построения комбинированного прогноза ансамбля моделей ARMA и ARIMA для полученных компонентов разложения.

Алгоритм разложения CEEMDAN основан на использовании процедуры EMD (*Empirical Mode Decomposition*). Процедура EMD декомпозиции временных рядов заключается разложении временного ряда в функции внутренних мод IMF (*Intrinsic Mode Function*). Функции IMF удовлетворяют следующим свойствам [3]:

1. Количество локальных минимумов и максимумов функций IMF должно быть равно количеству пересечений функцией нуля, либо отличаться не больше чем на единицу.
2. В любой точке среднее значение между огибающей линией по локальным минимумам и огибающей линией по локальным максимумам должно быть равно нулю.

Такие свойства функций IMF обеспечивают им в отдельности наличие более сильной автокорреляции в сравнении с исходным временным рядом, что гарантирует высокую эффективность прогнозов, основанных на данном свойстве. За счет этого совокупный прогноз по компонентам разложения является более точным, чем прогноз по исходному временному ряду.

Алгоритм декомпозиции EDM является итеративным и включает в себя следующие шаги [3]:

1. Определяются локальные минимумы и максимумы анализируемого временного ряда. Они соединяются линией кубического сплайна, образуя 2 огибающие линии, для которых находится средняя линия m_1 .
2. Определяется первый компонент процедуры “просеивания”: $h_{01} = x - m_1$.
3. Процесс итеративно повторяется пока k -ый компонент не станет функцией IMF: $h_{1k} = h_{1(k-1)} - m_{1k}$.
4. k -ый компонент процесса “просеивания” становится компонентом разложения: $c_1 = h_{1k}$.

5. Определяется ряд остатков: $r_1 = x - c_1$.
6. Процесс повторяется начиная с первого пункта для ряда остатков r_1 .

Такой алгоритм позволяет разложить исходный временной ряд в компоненты разных порядков частот, при этом сохраняя его изначальную структуру. Подхода EMD обладает некоторыми проблемами [4]: в рамках одного компонента могут присутствовать осцилляции существенно различающихся амплитуд, в свою очередь подобные осцилляции могут наблюдаться в различных компонентах, что называется “смешением мод” (*mode mixing*). Описанные проблемы декомпозиции могут быть решены с помощью построения ансамбля алгоритмов разложения и добавления адаптивного шума в начале каждой итерации к рассматриваемому временному ряду. Такой алгоритм называется SEEMDAN, особенности его работы представлены в работах [4, 5].

По результатам работы алгоритма временной ряд однопериодных приращений курса может быть представлен как сумма компонентов разложения и ряда остатков:

$$x = \sum_{n=1}^N c_i + r_N.$$

Второй этап предложенного метода заключается в построения прогноза с помощью ансамбля моделей ARMA-ARIMA. Полученные компоненты и ряд остатков по отдельности моделируются. Как было замечено ранее, компоненты IMF в общем случае могут быть нестационарными, хотя на практике в большинстве случаев нестационарными являются лишь низкочастотные компоненты. В качестве критерия стационарности используется статистический тест стационарности временных рядов ADF. Стационарные компоненты моделируются с помощью модели ARMA, нестационарные компоненты моделируются с помощью модели ARIMA.

С помощью каждой из моделей строится прогноз на s шагов вперед, результаты прогнозирования для всех компонентов и ряда остатков складываются, образуя прогноз для исходного временного ряда.

ПРОГНОЗИРОВАНИЕ ФОНДОВОГО РЫНКА И РЫНКА КРИПТОВАЛЮТ

Следуя предложенному методу, был проведен анализ однодневных приращений фондовых индексов DJIA (Dow Jones Industrial Average), RUT (Russell 2000), IXIC (NASDAQ Composite), а также криптовалюты BTC (Биткоин). Был построен пятишаговый прогноз, контрольный прогноз

строился с помощью модели ARMA. Качество прогноза оценивалось с помощью метрики MSE (*Mean Squared Error*), которая рассчитывается как среднеквадратичное отклонение прогнозных значений от реальных. Значения метрики MSE для прогнозов каждого из методов для каждого из рассматриваемых временных рядов приведены в таблице.

Значения метрики MSE для прогнозов каждого из методов для фондовых индексов и биткоина

	DJIA	RUT	IXIC	BTC
Ансамбль моделей	173.65	61.74	123.67	569976.84
ARMA	183.13	103.71	148.05	728738.03

По результатам экспериментов разработанный метод оказался точнее в каждом из случаев, что подтверждает его эффективность.

Библиографические ссылки

1. *Малюгин, В.И.* Рынок ценных бумаг: Количественные методы анализа / *В.И. Малюгин*. Москва: Дело, 2003. 320 с.
2. *Pin, L.* Stock Index Prediction Based on Time Series Decomposition and Hybrid Model / *P. Lv, Q. Wu, J. Xu, Y. Shu* // *Entropy*. 2022. v. 24. p. 146-163.
3. *Huang, N.E.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis / *N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N. Yen, C. Tung, H.H. Liu* // *Proc. The Royal Society*. 1998. v. 454. p. 903-995.
4. *Torres, M.E.* A complete ensemble empirical mode decomposition with adaptive noise. / *Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P.* // *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2011. p. 4144–4147.
5. *Wu, Z.* Ensemble empirical mode decomposition: A noise-assisted data analysis method / *Z. Wu, N.E. Huang* // *Advances in Adaptive Data Analysis*. 2009. v. 1. №. 1. p. 1–41.