

DOI: 10.17323/2587-814X.2023.3.87.100

Прогнозирование финансовых временных рядов с использованием сингулярного спектрального анализа*

А.В. Зиненко

E-mail: anna-z@mail.ru

Сибирский федеральный университет

Адрес: Россия, 660041, г. Красноярск, Свободный проспект, д. 79

Аннотация

Финансовые временные ряды представляют собой объемные массивы информации по котировкам и объемам торгов акций, валют и других биржевых и внебиржевых инструментов. Анализ и прогнозирование таких рядов всегда представляли особый интерес как для исследователей-аналитиков, так и для инвесторов-практиков. Однако, финансовые временные ряды имеют свою специфику, не позволяющую найти единственно верный и работающий метод прогнозирования. В настоящее время алгоритмы машинного обучения позволяют анализировать большие объемы данных и производить тестирование полученных моделей. Современные технологии позволяют тестировать и применять сложные методы прогнозирования, требующие объемных вычислений. Они дают возможность развивать математическую базу прогнозирования, комбинировать различные подходы в одном методе. Примером такого современного подхода является метод сингулярного спектрального анализа (SSA), который сочетает в себе разложение временного ряда в сумму временных рядов, метод главных компонент и рекуррентное прогнозирование. Целью настоящей работы является анализ возможности применения SSA метода к финансовым временным рядам. Метод SSA был рассмотрен в сравнении с другими распространенными методами прогнозирования финансовых временных рядов: ARIMA, разложение Фурье и рекуррентная нейронная сеть. Для реализации методов был разработан программный алгоритм на языке Python. Также была осуществлена апробация метода на временных рядах котировок российских и американских акций, валют и криптовалют.

Ключевые слова: нестационарные временные ряды, прогнозирование, сингулярный спектральный анализ, метрики ошибок

Цитирование: Зиненко А.В. Прогнозирование финансовых временных рядов с использованием сингулярного спектрального анализа // Бизнес-информатика. 2023. Т. 17. № 3. С. 87–100. DOI: 10.17323/2587-814X.2023.3.87.100

* Статья опубликована при поддержке Программы НИУ ВШЭ «Университетское партнерство»

Введение

К финансовым временным рядам можно отнести наборы данных по рыночным ценам различных финансовых инструментов, таких как ценные бумаги, валюты, производные контракты и драгоценные металлы. Такого рода временные ряды имеют некоторые особенности, которые следует учитывать при применении к ним тех или иных методов прогнозирования. Базовые инвестиционные теории, такие как модели Марковитца и Шарпа или модель ценообразования опционов Блэка-Шоулза [1], предполагали, что рыночные временные ряды являются случайными и подчиняются закону нормального распределения. Однако параллельно проводимые исследования, к примеру, Бенуа Мандельбротом [1, 2] и Эдгаром Петерсом [3, 4] доказали, что ценовые временные ряды неслучайны и, соответственно, нестационарны, поэтому применение к их прогнозу методов, основанных на предположении о случайности процессов, не даст адекватного прогноза.

Можно выделить следующие особенности финансовых временных рядов.

1. Как было указано выше, временные ряды рыночных цен нестационарны. Это значит, что среднее и дисперсия непостоянны на протяжении исследуемого промежутка. Поэтому прежде, чем применять к таким временным рядам методы, которые хорошо работают на случайных процессах, следует привести их к стационарному виду. Это можно сделать, к примеру, взятием разниц, как предлагает модель ARIMA.
2. Финансовые временные ряды персистентны. Это означает, что последующие показатели сильно зависят от предыдущих. К этому близка характерная для хаоса чувствительность к начальным условиям. Определить, персистентен ли временной ряд, позволяет показатель Херста [3]. Если данный показатель колеблется в пределах от 0,5 до 1, то временной ряд является персистентным. Херст называл такого рода процессы «процессами с долговременной памятью». Расчет показателя Херста для много численных временных рядов подтверждает их персистентный характер [4, 5].
3. По рыночным временным рядам имеется в открытом доступе большой объем информации. Это делает возможным применять к ним методы машинного обучения и анализа данных, в том числе и глубокое обучение, которое требует больших объемов тренировочной и тестовой выборки. Аналитики и трейдеры с успехом используют алгоритмы торгов, основанные на искусственных нейронных сетях и других методах машинного обучения. Финансовые данные доступны в разных временных промежутках, от минутных до недельных. С одной стороны, если брать минутные данные за несколько лет, то можно получить очень большой объем для обучения нейросети. Но с другой стороны, такая выборка может быть сильно зашумленной [6].
4. Финансовые временные ряды недифференцируемы, но при этом непрерывны. Если посмотреть на кривую биржевых котировок, то можно увидеть, что она не является гладкой. Это означает невозможность провести к ней касательную и, следовательно, невозможность посчитать производную в любой точке. Примером недифференцируемой непрерывной функции является функция Вейерштрасса [1], которая напоминает нетрендовый биржевой ряд. Данную специфику следует учитывать при применении определенных методов, например, анализа Фурье, которые разлагает временной ряд в сумму тригонометрических функций. В случае применения анализа Фурье к недифференцируемым рядам следует сгладить исходный ряд, например, скользящим средним.

Все перечисленные особенности следует учитывать при выборе методов прогнозирования финансовых временных рядов. К примеру, долговременную память могут учитывать некоторые виды рекуррентных нейросетей, такие как LSTM нейросети, способные в процессе обучения выбирать из прошлых данных те, которые наиболее существенно влияют на прогнозируемые значения.

Для проверки адекватности прогноза и сравнения методов прогнозирования следует разделять временной ряд на тренировочную и тестовую составляющую. На тренировочной составляющей алгоритм обучается (например, подбираются параметры модели Arima, коэффициенты Фурье, коэффициенты регрессии и т.д.). Затем полученная на тренировочном временном ряду модель применяется для прогноза, и прогнозный временной ряд сравнивается с тестовым с использованием метрик [7].

Все внутренние метрики основаны на оценке разницы между фактическими и прогнозными значениями. Средняя абсолютная ошибка (MAE) показывает среднее арифметическое разниц между фактическими и прогнозными значениями, средняя квадратичная ошибка (MSE) – среднее

арифметическое квадратов разниц, а средняя абсолютная ошибка в процентах (MAPE) – среднее арифметическое относительных разниц. MAPE вычисляется по следующей формуле

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \cdot 100 \right|, \quad (1)$$

где n – число наблюдений в тестовой выборке;

y_i – фактическое значение показателя в тестовой выборке;

\hat{y}_i – предсказанное значение показателя.

В тестировании предложенного метода прогнозирования была использована метрика MAPE, так как она позволяет оценить отклонение прогнозных данных от фактических в относительном выражении, то есть показывает отклонение прогноза от факта в процентах.

1. Материалы и методы

В данной работе прогноз финансовых временных рядов был осуществлен с использованием сингулярного спектрального анализа (SSA) или метода гусеницы. Как и анализ Фурье, SSA раскладывает исходный временной ряд на сумму компонент. Однако, в анализе Фурье временные ряды – слагаемые представляют собой периодические функции разной частоты и амплитуды, тогда как SSA разделяет исходный ряд на трендовую, периодическую и шумовую составляющие [8, 9]. Таким образом, SSA учитывает такие особенности финансовых временных рядов, как нестационарность и недифференцируемость.

Метод гусеницы – это вариант SSA анализа, разработанный в СССР в конце 80-х годов, к которому отечественные ученые пришли независимо от зарубежных коллег. В настоящее время наиболее глубоко раскрывает данный метод работа Н.Э. Голяндиной «Метод «Гусеница» – SSA анализ временных рядов» [10]. Отметим, что и зарубежные авторы ссылаются на работу Голяндиной как на первоисточник [11, 12]. Другие работы, посвященные SSA анализу, также принадлежат российским авторам: Л.Н. Леонтьевой [13], Д.Л. Данилову [14]. Авторы работ и пособий выделяют две проблемы при реализации метода: выбор длины гусеницы и выбор главных компонент. Если вторую проблему можно решить, рассмотрев вклад каждой компоненты в общую дисперсию, то по выбору длины гусеницы рекомендации носят скорее эвристический характер [15].

Суть метода SSA состоит в том, что исходный временной ряд трансформируется в матрицу, затем матрица с помощью сингулярного разложения разделяется на компоненты (здесь используется метод главных компонент). Следующий шаг зависит от целей анализа: либо компоненты разделяются на трендовую, периодическую и шумовую составляющие, и по ним производится анализ и прогноз (шумовые компоненты убираются), либо выбираются главные компоненты, и по ним строится продолжение ряда на длину заданного при формировании исходной матрицы шага. В данной работе использован второй подход. Алгоритм метода SSA выглядит следующим образом.

Анализируемый временной ряд имеет длину n . Выбирается длина гусеницы L , $2 \leq L \leq n/2$ и строится матрица X , полученная сдвигом каждой следующей колонки на одно значение размерности $L \times (n - L + 1)$.

$$\begin{matrix} x_1 & x_2 & \cdots & \cdots & x_{n-L+1} \\ x_2 & x_3 & \cdots & \cdots & x_{n-L+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_L & x_{L+1} & \cdots & \cdots & x_n \end{matrix} \quad (2)$$

Затем строится сингулярное разложение матрицы X :

$$X = U \sigma V^T, \quad (3)$$

где U – левый сингулярный вектор;

σ – диагональная матрица значений;

V^T – правый сингулярный вектор.

Полученные элементы векторов сортируются по убыванию собственных значений. Исходная матрица X разлагается на L элементарных матриц:

$$X_j = \sum_{j=1}^d \sigma_j U_j V_j^T, \quad j = \overline{1, L}, \quad (4)$$

где σ_j – j -е собственное значение;

$U_j V_j^T$ – соответствующие ему левый и правый сингулярный вектор;

d – ранг матрицы X .

Матрицы X_j имеют размерность $L \times (n - L + 1)$. Для того, чтобы привести данные матрицы к одномерному виду используется диагональное усреднение. Каждая строка матрицы сдвигается на $i = 1 \dots n - L + 1$ значение, затем рассчитываются средние значения по столбцу, которые и составляют расчетные значения j -й компоненты исходного временного ряда. Процедура диагонального усреднения показана формулой (5).

токорреляции уровней ряда, в котором по оси X называются временные лаги, а по оси Y — значения коэффициента корреляции между соответствующими лагу уровнями. Порядок авторегрессии выбирается равным такому временному лагу, при котором коэффициент корреляции принимает последнее максимальное отличное от нуля значение. Аналогичным образом выбирается порядок скользящего среднего, только вместо коэффициентов автокорреляции рассчитываются коэффициенты частичной автокорреляции. Частичная автокорреляция отличается от автокорреляции тем, что не учитывает влияние уровней, находящихся между текущим уровнем и уровнем, отстающим на временной лаг. Очевидно, что при единичном лаге автокорреляция и частичная автокорреляция совпадают.

Для осуществления прогноза по модели ARIMA на Python достаточно определить параметры p , d , q модели. Формально модель ARIMA описывается следующим образом

$$\Delta^d y_t = \sum_{i=1}^p \alpha_i \Delta^d y_{t-i} - \sum_{j=1}^q \beta_j \varepsilon_{t-j}, \quad (9)$$

где Δ^d — разность d -го порядка, необходимая для достижения стационарности;

α_i — коэффициенты авторегрессии p -го порядка;

β_j — коэффициенты скользящего среднего q -го порядка;

ε_{t-j} — ошибки предсказания скользящего среднего.

Коэффициенты α_i и β_j оцениваются и подставляются в прогноз.

Преобразование Фурье разлагает периодическую функцию на сумму синусоид и косинусоид с известными частотами, амплитудами и фазами. В общем виде разложение Фурье показано формулой (9).

$$y_t = \bar{y}_t + \sum_{i=1}^{N/2} (\alpha_i \cos \omega_i t + b_i \sin \omega_i t), \quad (10)$$

где y_t — преобразованное значение временного ряда;

\bar{y}_t — среднее значение исходного временного ряда;

ω_i — частота i -й гармоники (первая частота соответствует периоду функции, остальные кратны ей);

α_i, β_i — коэффициенты, которые необходимо определить.

Коэффициенты ряда Фурье вычисляются по следующим формулам:

$$a_1 = \frac{2}{N} \sum_{t=0}^{2\pi(N-1)/N} y_t \cos t, \quad b_1 = \frac{2}{N} \sum_{t=0}^{2\pi(N-1)/N} y_t \sin t \quad (11)$$

для первой гармоники,

$$a_2 = \frac{2}{N} \sum_{t=0}^{2\pi(N-1)/N} y_t \cos 2t, \quad b_2 = \frac{2}{N} \sum_{t=0}^{2\pi(N-1)/N} y_t \sin 2t \quad (12)$$

для второй гармоники и так далее. Обычно для прогноза достаточно первых двух–трех гармоник. Коэффициенты Фурье, полученные по формулам (11) и (12), подставляются в уравнение (10), и по данному уравнению осуществляется прогноз. В данной работе для прогнозирования методом разложения Фурье был использован код на языке программирования Python, разработанный автором.

Алгоритм прогнозирования временных рядов с использованием **рекуррентной нейросети** можно представить следующим образом.

1. Исходный временной ряд длиной n преобразовывается в матрицу, в которой строки соответствуют шагу L для построения прогноза. Массив размерностью $(L \times (n - L + 1))$ подается на вход в рекуррентный слой. Также задается m — шаг, на сколько периодов вперед строится прогноз.
2. Внутри рекуррентного слоя поданный массив обрабатывается активационной функцией, на выход подается массив размерности, заданной пользователем. Если данный слой последний, то размерность выходного массива равна $(m \times 1)$.
3. Готовое решение сравнивается с фактическими данными. Задается функция потерь (разница между фактическими и предсказанными нейросетью данными), и с помощью оптимизационной функции реализуется алгоритм обратного распространения ошибки. Алгоритм обратного распространения ошибки изменяет заданные случайным образом на этапе прямой работы нейросети веса таким образом, чтобы минимизировать функцию потерь. Большие массивы данных для «подгонки» весов разделяются на пакеты (батчи), то есть оптимизатор меняет веса не после подачи каждого сигнала, а после подачи пакета. Размер пакета задается при конструировании нейросети и обычно является степенью двойки. Число эпох определяет количество «прогонок» нейросети для успешного обучения. Качество обучения определяет функция потерь на тренировочных данных и метрика ошибки на тестовых.

2. Результаты и обсуждение

Для практического применения SSA был разработан алгоритм на языке Python с использованием библиотек numpy, pandas, matplotlib, sclearn. Данные по зарубежным финансовым временным рядам

были получены с сайта yahoo.finance при помощи библиотеки Python yfinance, а данные по котировкам отечественных акций – при помощи библиотеки Python arimoex.

В целях анализа были выбраны дневные котировки топовых российских и американских акций за период с июня 2022 по март 2023. Всего было проанализировано 30 компаний – 15 российских и 15 американских. Данные за 110 дней были разделены на тренировочную и тестовую выборку. Тренировочная выборка включала в себя 100 значений, тестовая – 10. Поскольку прогноз осуществлялся на период L – длину гусеницы, то, соответственно, параметр L был выбран равный 10. Таким образом, исходная матрица X имела размерность 10×91 , и в результате сингулярного разложения была разложена на 10 матриц такой же размерности. Данные матрицы были преобразованы в одномерные массивы размерностью 100, и из них были выбраны те, собственные значения матриц которых дают наибольший вклад в дисперсию. Затем полученные прогнозные значения сравнивались с тестовой выборкой с помощью метрики MAPE. В таблице 1 показаны значения MAPE SSA-прогноза анализируемых акций.

В среднем ошибка прогноза составила 5,54%, в том числе по отечественным акциям – 4,62%, а по акциям американских компаний – 6,46%. Отметим,

что присутствует сильный выброс – по компании JPMorgan Chase & Co. ошибка оказалось значительной – 26%. По российским компаниям достаточно большая ошибка только по компании Русал – 11% (это вполне приемлемый для прогноза результат). По акциям компаний США существенная ошибка также по компании Mastercard – 17%. Также достаточно велики ошибки по компаниям United Health (9,5%) и Advanced Micro Devices (11%). Самые лучшие прогнозы были осуществлены по акциям Норильский никель (ошибка – 0,5%), МТС (ошибка – 1,7%) и McDonald's (ошибка – 1,2%). На рисунке 2 показаны несколько графиков, иллюстрирующих SSA метод, при этом визуализированы как «плохие», так и «хорошие» прогнозы.

Для сравнения SSA анализа с методами ARIMA, разложение Фурье и рекуррентная нейронная сеть был осуществлен прогноз по тем же группам акций за тот же временной промежуток. Параметры p и d модели ARIMA для всех акций составили соответственно 2 и 1, тогда как параметр q менялся в зависимости от графика автокорреляции. Разложение Фурье осуществлялось по трем гармоникам. Рекуррентная нейронная сеть была сконструирована с использованием библиотек Keras и Tensor Flow Python. Поскольку для нейросети анализируемый промежуток весьма мал, во избежание переобучения был выбран один

Таблица 1.

MAPE метода SSA по акциям

Россия	MAPE	США	MAPE
Полиметал	7,00%	Amazon	5,00%
Полюс Золото	5,30%	Apple	4,70%
Мечел	7,80%	American Express Company	6,20%
ГМК «Норильский никель»	0,50%	Tesla	2,60%
Яндекс	5,20%	Advanced Micro Devices	11,00%
Аэрофлот	2,90%	Pfizer	2,10%
ВТБ	4,60%	Netflix	2,12%
Магнит	2,90%	Microsoft	2,00%
Алроса	5,20%	Mastercard	17,00%
Тинькофф Груп	6,10%	Visa	2,10%
Русал	11,00%	Starbucks	2,30%
Новатэк	2,40%	JPMorgan Chase & Co.	26,00%
Сургутнефтегаз	3,20%	McDonald's	1,20%
МТС	1,70%	Boeing	3,10%
Северсталь	3,50%	United Health	9,50%

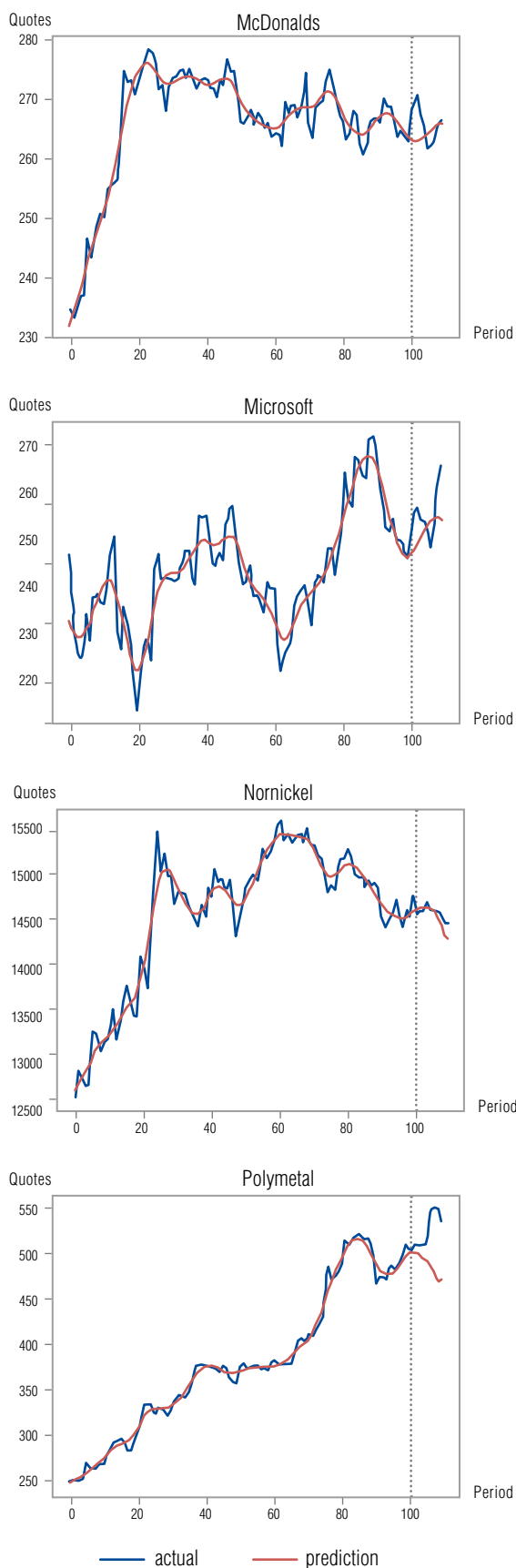


Рис. 2. Визуализация прогноза SSA методом.

Таблица 2.

**MAPE методов ARIMA,
разложение Фурье, RNN**

Россия	MAPE, ARIMA	MAPE, Фурье	MAPE, RNN
Полиметал	1,70%	27,60%	95,77%
Полус Золото	1,50%	17,00%	99,79%
Мечел	1,80%	22,20%	19,50%
ГМК «Норильский никель»	0,50%	1,50%	54,83%
Яндекс	1,20%	3,20%	95,51%
Аэрофлот	0,70%	8,80%	5,86%
ВТБ	0,80%	1,50%	3,60%
Магнит	0,40%	3,30%	99,05%
Алроса	0,70%	2,30%	29,19%
Тинькофф Груп	1,00%	3,90%	98,34%
Русал	0,70%	5,00%	2,65%
Новатэк	0,80%	2,60%	95,85%
Сургутнефтегаз	0,80%	3,50%	2,19%
МТС	1,00%	8,40%	81,83%
Северсталь	1,10%	18,40%	95,12%
Среднее по акциям РФ	0,98%	8,61%	58,61%
Amazon	1,50%	3,10%	56,39%
Apple	1,60%	6,40%	69,55%
American Express Company	2,30%	8,50%	70,46%
Tesla	3,10%	5,90%	76,90%
Advanced Micro Devices	2,00%	17,00%	34,24%
Pfizer	1,00%	17,10%	8,81%
Netflix	1,70%	3,00%	83,86%
Microsoft	1,30%	5,90%	81,80%
Mastercard	1,40%	3,80%	86,55%
Visa	1,30%	4,80%	78,09%
Starbucks	1,10%	3,50%	53,21%
JPMorgan Chase & Co.	2,70%	4,10%	64,73%
McDonald's	0,80%	2,00%	83,24%
Boeing	2,10%	10,80%	72,31%
United Health	0,70%	9,40%	91,70%
Среднее по акциям США	1,64%	7,18%	67,46%
Общее среднее	1,31%	7,90%	63,03%

RNN слой. Активационная функция, функция потерь и функция-оптимизатор были выбраны, исходя из рекомендованных в источнике [6] параметров для прогнозирующих рекуррентных нейронных сетей. Размер пакета был выбран минимальный – 2, а количество эпох было выбрано такое, при котором функция потерь и ошибка (средняя абсолютная ошибка и средняя абсолютная ошибка в процентах соответственно) перестали меняться. В большей части экспериментов использовалось 20 эпох. Прогноз, так же, как и SSA методом, осуществлялся на 10 дней вперед.

Как видно из *таблицы 2*, метод ARIMA по точности прогноза значительно превзошел прочие методы, включая SSA. Однако SSA анализ показал более точные результаты, чем разложение Фурье и рекуррентная нейронная сеть, а по акциям Норильский

Никель, МТС и McDonald's вполне сопоставим с ARIMA. Разложение Фурье в целом показало на акциях также неплохие результаты, выбросы ошибки присутствовали только по акциям Мечел, Advanced Micro Devices и Pfizer. Что касается рекуррентной нейронной сети, то на коротком промежутке по акциям она не работает, что подтвердили неприемлемые размеры ошибок.

Далее сингулярный спектральный анализ в сравнении с прочими методами был применен к валютному рынку и рынку криптовалют, которые являются более динамичными, чем рынок акций. Временной промежуток остался тем же – с июня 2022 по март 2023. Были взяты десять валютных пар и десять котировок криптовалют к доллару США. Средняя абсолютная ошибка в процентах по сравниваемым методам показана в *таблице 3*.

Таблица 3.

MAPE методов SSA, ARIMA, преобразования Фурье и RNN по валютам

Валюта	MAPE, SSA	MAPE, ARIMA	MAPE, Фурье	MAPE, RNN
Евро / Доллар США	59,00%	0,50%	2,70%	1,89%
Фунт / Доллар США	0,70%	0,60%	2,10%	2,22%
Доллар США / Юань	3,70%	0,50%	1,70%	1,47%
Доллар США / Рубль	37,90%	0,30%	12,20%	30,46%
Доллар США / Иена	40,20%	0,60%	1,80%	69,06%
Доллар США / Гонконгский доллар	43,60%	0,00%	1,80%	0,33
Доллар США / Южноафриканский Рэнд	8,30%	0,80%	5,00%	2,27%
Австралийский Доллар / Доллар США	0,90%	0,60%	1,90%	0,33%
Доллар США / Мексиканское Песо	1,70%	1,10%	5,20%	1,36%
Новозеландский Доллар / Доллар США	2,00%	0,70%	1,90%	2,42%
Среднее по фиатным валютам	16,38%	0,57%	3,63%	15,90%
Bitcoin	6,80%	3,10%	9,80%	99,90%
Ethereum	8,20%	3,10%	8,30%	98,75%
Binance Coin	47,3%	1,80%	4,00%	6,57%
Polygon	8,10%	4,00%	8,00%	6,25%
Lightcoin	12,30%	5,70%	8,40%	74,24%
Ripple	1,90%	1,90%	2,50%	6,29%
Polkadot	6,40%	3,50%	4,10%	5,70%
Chainlink	7,40%	3,30%	4,60%	9,44%
Avalanche	17,40%	4,70%	4,8%	11,13%
Tron	45,50%	4,00%	6,70%	4,52%
Среднее по криптовалютам	16,13%	3,51%	6,08%	34,82%
Итоговое среднее	16,26%	2,04%	4,85%	25,36%

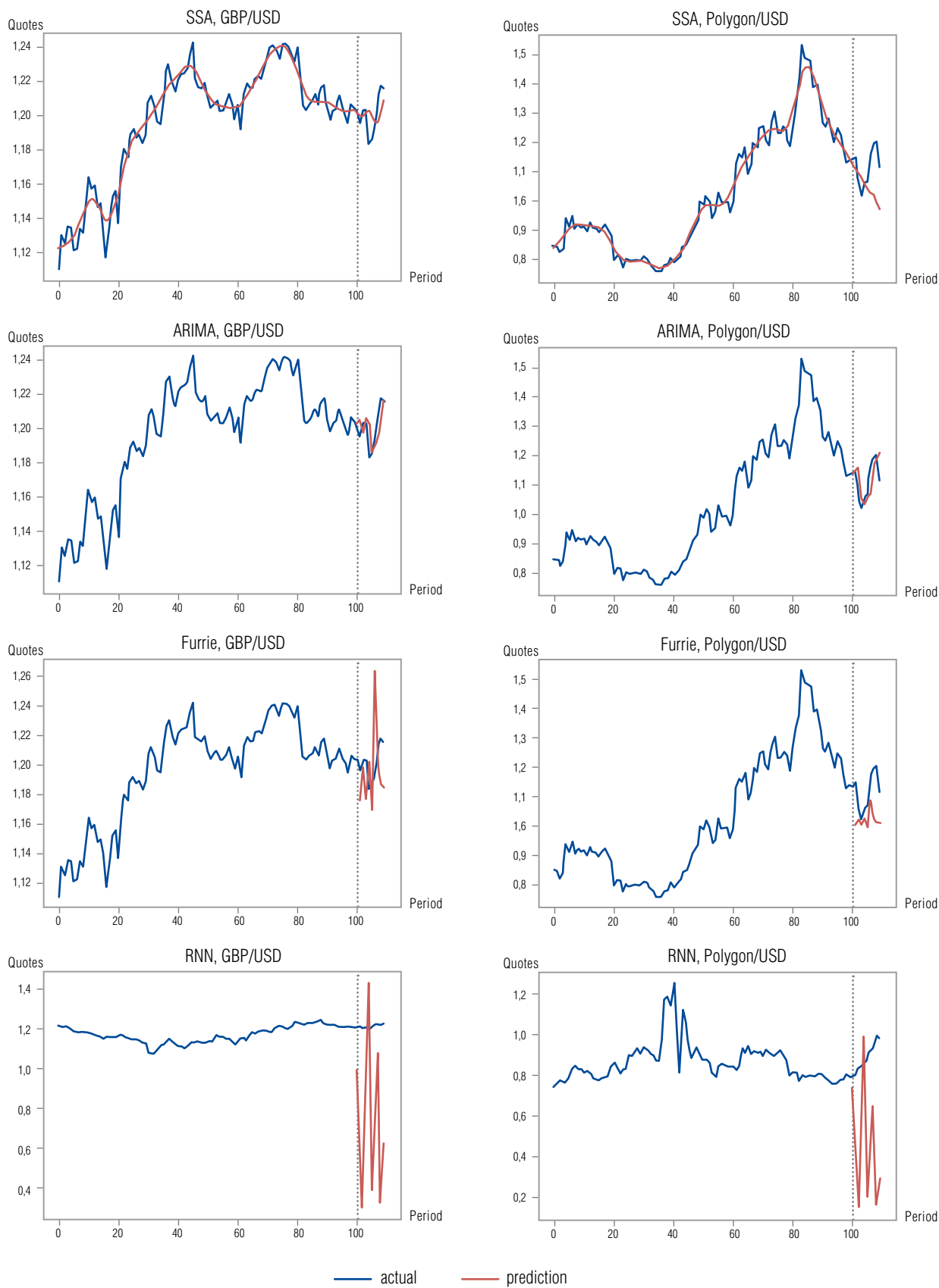


Рис. 3. Визуализация прогнозов по валютам методами SSA, ARIMA, разложение Фурье и RNN.

С валютами, как фиатными, там и крипто, метод SSA показал не очень хорошие результаты. Если по криптовалютам выброс произошел только по валютам BNB и Tron, то по фиатным валютным парам метод показал крупную ошибку в четырех случаях из десяти. Метод ARIMA на валютах работал так же хорошо, как и на акциях, а методы разложения Фурье и рекуррентная нейросеть существенно улучшили точность прогноза. При этом прогноз рекуррентной нейросети остался неудовлетворительным. Отметим, что все рассмотренные методы на криптовалютах работают хуже, чем на фиатных. Это может быть связано с тем, что в силу молодости крипторынка на нем все еще остаются лазейки для арбитражеров. На *рисунке 3* показаны графики прогнозов некоторых фиатных и криптовалют рассмотренными методами.

В качестве решения проблемы плохого прогноза SSA как на фиатных, так и на криптовалютах, мы увеличили временной промежуток вдвое. Новые тренировочная и тестовая выборки по тем же валютам составили 200 и 20 (длина гусеницы соответственно была так же выбрана 20). Средняя абсолютная ошибка в процентах при новых условиях показана в *таблице 4*.

При увеличении временного промежутка результаты прогноза методом SSA значительно улучшились, тогда как точность прогноза методом ARIMA практически не изменилась, а методом Фурье даже ухудшилась. Рекуррентная нейронная сеть существенно улучшила точность прогноза, однако по двум фиатным валютам и трем криптовалютам остались значительные выбросы. Лучшая точность осталась за моделью ARIMA. На *рисунке 4* показаны прогнозы по валютам на длинном временном промежутке.

Таблица 4.

MAPE методов SSA и рекуррентной нейросети по валютам (длинный промежуток)

Валюта	MAPE, SSA	MAPE, ARIMA	MAPE, Фурье	MAPE, RNN
Евро / Доллар США	2,60%	0,40%	3,00%	2,44%
Фунт / Доллар США	8,20%	0,50%	1,20%	2,62%
Доллар США / Юань	0,90%	0,40%	1,10%	2,40%
Доллар США / Рубль	1,50%	1,10%	16,30%	50,53%
Доллар США / Иена	6,50%	0,50%	1,50%	78,31%
Доллар США / Гонконгский доллар	1,00%	0,00%	0,90%	0,22%
Доллар США / Южноафриканский Рэнд	1,80%	0,60%	6,60%	2,94%
Австралийский Доллар / Доллар США	3,10%	0,50%	1,70%	0,38%
Доллар США / Мексиканское Песо	1,10%	0,70%	8,1%	1,53%
Новозеландский Доллар / Доллар США	3,20%	0,50%	1,10%	2,99%
Среднее по фиатным валютам	3,16%	0,52%	4,48%	14,44%
Bitcoin	9,90%	2,30%	13,30%	99,83%
Ethereum	48,80%	2,40%	9,90%	97,76%
Binance Coin	3,20%	1,50%	3,80%	7,22%
Polygon	16,50%	3,70%	20,50%	7,23%
Lightcoin	12,40%	3,90%	19,40%	50,07%
Ripple	4,40%	1,50%	6,10%	8,88%
Polkadot	19,70%	3,20%	5,80%	6,32%
Chainlink	10,80%	2,90%	5,20%	8,55%
Avalanche	7,40%	4,00%	6,2%	18,19%
Tron	7,00%	2,70%	9,80%	4,73%
Среднее по криптовалютам	14,01%	2,81%	10,02%	30,88%
Итоговое среднее	8,58%	1,67%	7,25%	22,66%

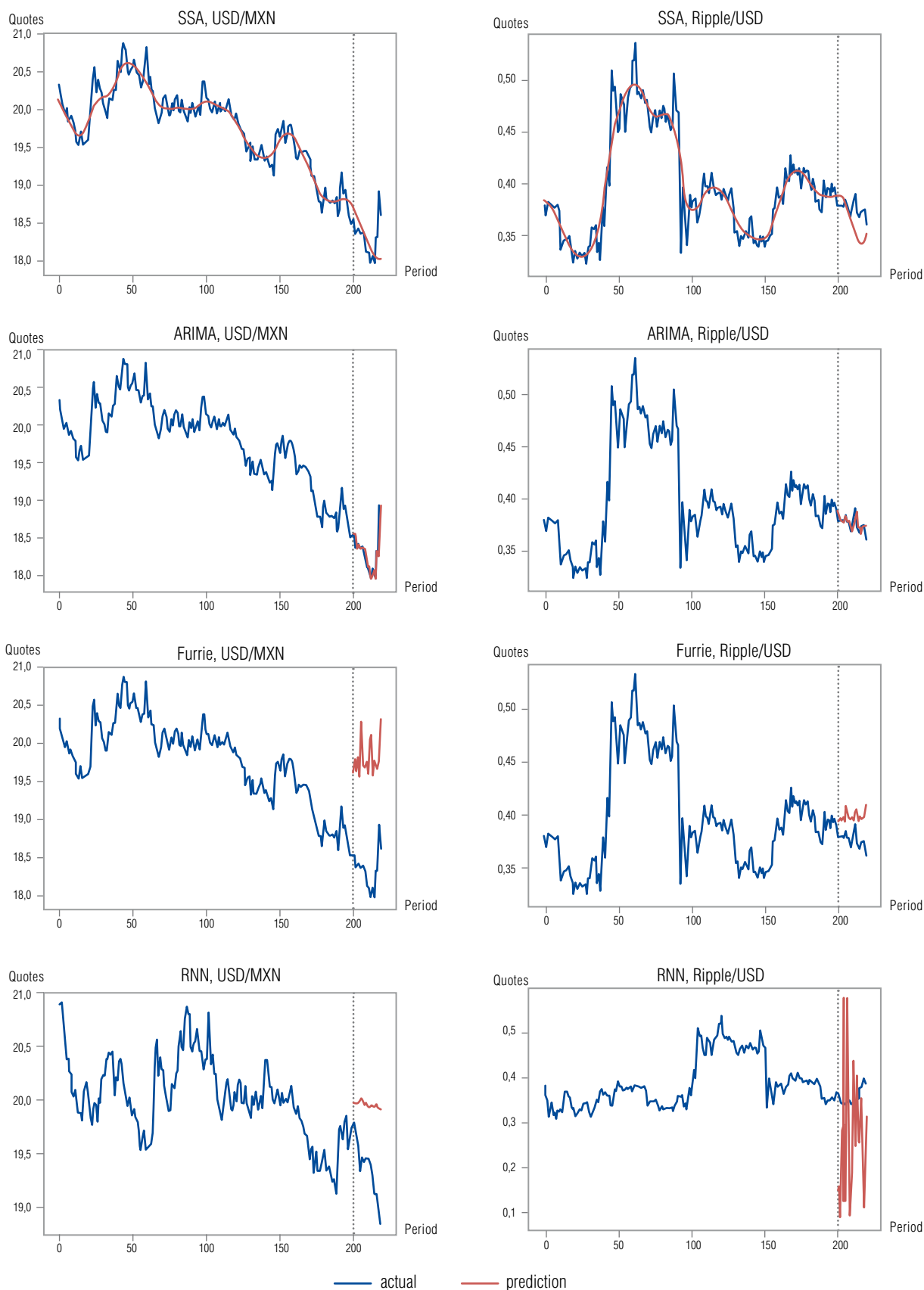


Рис. 4. Визуализация прогнозов по валютам методами SSA, ARIMA, разложение Фурье и RNN (длинный временной промежуток).

Заключение

В работе был рассмотрен метод прогнозирования временных рядов SSA анализ (метод гусеницы). Метод был реализован путем разработки алгоритма на языке Python, затем протестирован на 30 временных рядах котировок акций РФ и США, а также на 20 фиатных и криптовалютах по отношению к доллару США. Для сравнения были взяты три метода прогнозирования: ARIMA, разложение Фурье и рекуррентная нейронная сеть. Метод SSA анализа во всех случаях, кроме короткого временного промежутка по валютам, показал второй по точности после метода ARIMA результат. По некоторым бумагам и акциям ошибка SSA метода сравнима с ошибкой метода ARIMA. При этом увеличение временного промежутка значительно улучшило результаты применения SSA метода, тогда как результаты метода ARIMA остались неизменными.

Можно сделать вывод о том, что хоть метод SSA анализа показывает меньшую точность прогноза, чем общепринятый в анализе финансовых временных рядов метод ARIMA, он может быть применен как к акциям, так и к другим финансовым инструментам, даже к таким волатильным, как криптовалюты. Его можно использовать как подтверждение результатов метода ARIMA, а также отдельно как метод прогнозирования. Отметим, что для анализа большого массива акций метод SSA удобнее, чем ARIMA, так как ARIMA требует пересчета по крайней мере порядка скользящего среднего для каждого временного ряда, тогда как для прогноза методом SSA достаточно одной главной компоненты. Кроме того, если рассматривать другие компоненты сингулярного разложения, то можно сделать выводы о соотношении трендовости, периодичности и шума в анализируемом временном ряде, что не позволяет ни один из прочих рассмотренных методов. ■

Литература

1. Mandelbrot B., Fisher A., Calvet L. A multifractal model of asset returns // Yale Cowles Foundation for Research in Economics. 1997. Discussion Paper No. 1164.
2. Мандельброт Б. Фракталы, случай и финансы. Москва, Ижевск: НИЦ «Регулярная и хаотическая динамика», 2004.
3. Петерс Э. Фрактальный анализ финансовых рынков. Применение хаоса в инвестициях и экономике. М.: Интернет-трейдинг, 2004.
4. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка. М.: Мир, 2000.
5. Зиненко А.В. R/S анализ на фондовом рынке // Бизнес-информатика. 2012. № 3(21). С. 21–27.
6. Шоле Ф. Глубокое обучение на Python. СПб.: Питер, 2018.
7. Kong Q., Han J., Jin X., Li C., Wang T., Bai Q., Chen Y. Polar motion prediction using the combination of SSA and ARMA // Geodesy and Geodynamics. 2023. Vol. 14. No. 4. P. 368–376. <https://doi.org/10.1016/j.geog.2022.12.004>
8. Li K., Zhang Z., Guo H., Li W., Yan Y. Prediction method of pipe joint opening-closing deformation of immersed tunnel based on singular spectrum analysis and SSA-SVR // Applied Ocean Research. 2023. Vol. 135. 103526. <https://doi.org/10.1016/j.apor.2023.103526>
9. Montalvo C., Pantera L., Lipcsei S., Torres L.A. Signal processing applied in cortex project: From noise analysis to OMA and SSA methods. Annals of Nuclear Energy. 2022. Vol. 175. 109193. <https://doi.org/10.1016/j.anucene.2022.109193>
10. Голяндина Н.Э. Метод «Гусеница» – SSA: анализ временных рядов. СПб.: СПбГУ, 2004.
11. Coussin M. Singular spectrum analysis for real-time financial cycles measurement // Journal of International Money and Finance. 2022. Vol. 120. 102532. <https://doi.org/10.1016/j.jimonfin.2021.102532>
12. Lahmiri S., Bekiros S., Bezzina F. Evidence of the fractal market hypothesis in European industry sectors with the use of bootstrapped wavelet leaders singularity spectrum analysis // Chaos, Solitons & Fractals. 2022. Vol. 165. Part 1. 112813. <https://doi.org/10.1016/j.chaos.2022.112813>
13. Леонтьева Л.Н. Многомерная гусеница, выбор длины и числа компонент // Машинное обучение и анализ данных. 2011. № 1. С. 5–15.
14. Солнцев В.Н., Данилов Д.Л., Жиглявский А.А. Главные компоненты временных рядов: Метод «Гусеница». СПб.: СПбГУ, 1997.
15. Baharanchi S.A., Vali M., Modares M. Noise reduction of lung sounds based on singular spectrum analysis combined with discrete cosine transform // Applied Acoustics. 2022. Vol. 199. 109005. <https://doi.org/10.1016/j.apacoust.2022.109005>

Об авторе

Зиненко Анна Викторовна

к.т.н., доцент;

доцент, кафедра экономической и финансовой безопасности, Сибирский федеральный университет, 660041, г. Красноярск, Свободный проспект, д. 79;

E-mail: anna-z@mail.ru

Forecasting financial time series using singular spectrum analysis

Anna V. Zinenko

E-mail: anna-z@mail.ru

Siberian Federal University

Address: 79, Svobodny Prospect, Krasnoyarsk 660041, Russia

Abstract

Financial time series are big arrays of information on quotes and trading volumes of shares, currencies and other exchange and over-the-counter instruments. The analysis and forecasting of such series has always been of particular interest for both research analysts and practicing investors. However, financial time series have their own features, which do not allow one to choose the only correct and well-functioning forecasting method. Currently, machine-learning algorithms allow one to analyze large amounts of data and test the resulting models. Modern technologies enable testing and applying complex forecasting methods that require volumetric calculations. They make it possible to develop the mathematical basis of forecasting, to combine different approaches into a single method. An example of such a modern approach is the Singular Spectrum Analysis (SSA), which combines the decomposition of a time series into a sum of time series, principal component analysis and recurrent forecasting. The purpose of this work is to analyze the possibility of applying SSA to financial time series. The SSA method was considered in comparison with other common methods for forecasting financial time series: ARIMA, Fourier transform and recurrent neural network. To implement the methods, a software algorithm in the Python language was developed. The method was also tested on the time series of quotes of Russian and American stocks, currencies and cryptocurrencies.

Keywords: non-stationary time series, forecasting, singular spectrum analysis, error metrics

Citation: Zinenko A.V. (2023) Forecasting financial time series using singular spectrum analysis. *Business Informatics*, vol. 17, no. 3, pp. 87–100. DOI: 10.17323/2587-814X.2023.3.87.100

References

1. Mandelbrot B., Fisher A., Calvet L. (1997) A multifractal model of asset returns. *Yale Cowles Foundation for Research in Economics*, Discussion Paper No. 1164.
2. Mandelbrot B. (2004) *Fractals, case and finance*. Moscow, Izhevsk: Research Center “Regular and Chaotic Dynamics” (in Russian).
3. Peters E. (2004) *Fractal analysis of financial markets. Application of chaos in investment and economics*. Moscow: Internet Trading (in Russian).
4. Peters E. (2000) *Chaos and order in capital markets. A new analytical perspective on cycles, prices and market volatility*. Moscow: Mir (in Russian).
5. Zinenko A.V. (2012) R/S analysis in the stock market. *Business Informatics*, no. 3(21), pp. 21–27 (in Russian).
6. Cholet F. (2018) *Deep learning with Python*. Saint Petersburg: Peter (in Russian).
7. Kong Q., Han J., Jin X., Li C., Wang T., Bai Q., Chen Y. (2023) Polar motion prediction using the combination of SSA and ARMA. *Geodesy and Geodynamics*, vol. 14, no. 4, pp. 368–376. <https://doi.org/10.1016/j.geog.2022.12.004>
8. Li K., Zhang Z., Guo H., Li W., Yan Y. (2023) Prediction method of pipe joint opening-closing deformation of immersed tunnel based on singular spectrum analysis and SSA-SVR. *Applied Ocean Research*, vol. 135, 103526. <https://doi.org/10.1016/j.apor.2023.103526>
9. Montalvo C., Pantera L., Lipcsei S., Torres L.A. (2022) Signal processing applied in cortex project: From noise analysis to OMA and SSA methods. *Annals of Nuclear Energy*, vol. 175, 109193. <https://doi.org/10.1016/j.anucene.2022.109193>
10. Golyandina N.E. (2004) *Method “Caterpillar”-SSA: analysis of time series*. Saint Petersburg: St. Petersburg State University (in Russian).
11. Coussin M. (2022) Singular spectrum analysis for real-time financial cycles measurement. *Journal of International Money and Finance*, vol. 120, 102532. <https://doi.org/10.1016/j.jimonfin.2021.102532>
12. Lahmiri S., Bekiros S., Bezzina F. (2022) Evidence of the fractal market hypothesis in European industry sectors with the use of bootstrapped wavelet leaders singularity spectrum analysis. *Chaos, Solitons & Fractals*, vol. 165, part 1, 112813. <https://doi.org/10.1016/j.chaos.2022.112813>
13. Leontyeva L.N. (2011) Multidimensional caterpillar, choice of length and number of components. *Machine Learning and Data Analysis*, no. 1, pp. 5–15 (in Russian).
14. Solntsev V.N., Danilov D.L., Zhiglyavsky A.A. (1997) *Principal components of time series: Method “Caterpillar”*. Saint Petersburg: St. Petersburg State University (in Russian).
15. Baharanchi S.A., Vali M., Modares M. (2022) Noise reduction of lung sounds based on singular spectrum analysis combined with discrete cosine transform. *Applied Acoustics*, vol. 199, 109005. <https://doi.org/10.1016/j.apacoust.2022.109005>

About the author

Anna V. Zinenko

Cand. Sci. (Tech.), Associate Professor;

Associate Professor, Department of Economic and Financial Security, Siberian Federal University, 79, Svobodny Prospect, Krasnoyarsk 660041, Russia;

E-mail: anna-z@mail.ru