

Федеральное государственное автономное образовательное учреждение высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**
Факультет экономических наук

КУРСОВАЯ РАБОТА

**Прогнозирование динамики валютного курса евро/доллар на основе её
декомпозиции на несколько случайных процессов**

Название темы

по направлению подготовки Экономика
образовательная программа «Экономика и статистика»

Выполнил:

Студент группы БСТ184

Бернухова Анастасия Владимировна

Ф.И.О.

Руководитель:

преподаватель школы

финансов на ФЭН,

Гаращук Г. В.

степень, звание, должность Ф.И.О.

Москва 2021

Оглавление

Введение	3
Глава 1. Теоретический обзор	5
1.1. Обзор литературы	5
1.2. Описание метода «Гусеница»-SSA	7
Глава 2. Методология курсовой работы	11
Глава 3. Применение сингулярно-спектрального анализа в задаче отделения ряда от шума и прогнозирования	16
3.1. Описание выбранных данных	16
3.2. Применение сингулярно-спектрального анализа и группировка компонент	18
3.3. Применение модели ARIMA для прогнозирования динамики валютного курса	29
Выводы	36
Список использованной литературы	37

Введение

Данная курсовая работа посвящена исследованию динамики валютного курса евро/доллар и его прогнозированию с помощью сингулярно-спектрального анализа (SSA, Singular Spectrum Analysis) и методов машинного обучения, в частности, с использованием модели ARIMA¹, которая как раз предназначена для временных рядов.

Прогнозирование валютного курса является одной из важных проблем финансовой системы. Предсказание валютного курса позволяет экономическому агенту получить бóльшую прибыль при торговле на фондовом рынке. Проблема прогнозирования различных динамик, в частности, курса валют, остается актуальной также и из-за развития в настоящий момент методов анализа данных и машинного обучения.

Исследование различных подходов к проблеме позволяет продолжать поиски лучшего метода для ее решения, что делает данную работу актуальной несмотря на то, что на сегодняшний день уже существует множество различных работ, посвященных прогнозированию курса.

Исследуемая гипотеза состоит в том, что прогнозирование цены валютной пары после отделения временного ряда от шума будет качественнее, чем в случае нахождения предсказания без проделывания подобной обработки массива данных.

Объект исследования – валютный курс евро/доллар. Предмет – прогнозирование динамики данного курса.

Целью работы является проанализировать динамику валютного курса евро/доллар с точки зрения временного ряда: его разложения и группировки с помощью метода SSA, спрогнозировать значение валютного курса, используя методы машинного обучения, сравнить результаты при отделении ряда от шума и без данной процедуры.

Задачи исследования:

¹ ARIMA (autoregressive integrated moving average) – интегрированная модель авторегрессии скользящего среднего, или модель Бокса-Дженкинса [13]

- Изучить существующие работы, посвященные анализу динамики и прогнозирования различных показателей;
- Изучить метод сингулярно-спектрального анализа;
- Выбрать данные для анализа и применить к ним данный метод, используя Python;
- Проанализировать компоненты, полученные после разложения ряда;
- Выполнить визуальную группировку компонент и группировку на основе матрицы взвешенной корреляции;
- Проверить компоненты, полученные после группировок на нормальность для определения, какие из них являются шумом (колебаниями), а какие относятся к основным составляющим ряда (тренд, периодики), сгруппировать основные в одну компоненту, аналогично проделать и с шумовыми составляющими ряда;
- Применить модель ARIMA на очищенных от шума данных и на исходном временном ряде для прогнозирования валютного курса, сравнить результаты и качество моделей с помощью метрик качества.

Глава 1. Теоретический обзор

1.1. Обзор литературы

Благодаря развитию методов машинного обучения, с каждым днем появляется все больше и больше работ, посвященных анализу динамики цен на различные экономические блага, валютных курсов, различных макро- и микроэкономических показателей.

Сделаем некоторый обзор на существующие работы, в основе которых лежит идея о предсказании значений каких-либо переменных.

В исследовании Кристиана Л. Дуниса и Марка Уильямса [9] изучается применение моделей NNR² в задаче прогнозирования курса евро/доллар. Выбранные модели сравниваются с другими существующими способами поиска предсказания и оцениваются, помимо общеизвестных метрик качества, таких, как, например, среднеквадратическая ошибка, с помощью различных финансовых критериев. Модель NNR является неким обобщением многих других моделей, что в условиях отсутствия точных знаний о подходящей модели, безусловно, дает ей некоторое преимущество перед другими моделями. С другой стороны, такая обобщенность в то же время является и недостатком данной модели. Кроме того, для данной модели характерно длительное время, затрачиваемое на ее обучение, необходимость задания большого числа параметров и высокий риск переобучения. Все это может сильно усложнять анализ динамики курса валют.

Как было сказано ранее, в работах можно встретить различные модели прогнозирования. Например, в работе «Сравнение финансовых моделей для прогнозирования курса акций» («Comparison of Financial Models for Stock Price Prediction») [14] сопоставляются три модели прогнозирования: ARIMA, ANNs³ и стохастические модели. Вывод работы – последние из вышеназванных моделей обладают большей предсказательной способностью.

² NNR (neural networks regressions) – модели регрессии нейронных сетей

³ ANNs (artificial neural networks) – искусственная нейронная сеть

Итак, существует множество работ, посвященных задаче прогнозирования различных переменных, различающихся используемыми методами, моделями и поставленными задачами. Моделей прогнозирования существует довольно много, так, например, в статье Хоссейна Хассани и др. [10] сравниваются целых 17 методов прогнозирования, среди которых: ARIMA, ETS⁴, TBATS⁵, ARFIMA⁶, AR⁷, VAR⁸, BAR⁹, BVAR¹⁰; Random Walk¹¹.

Еще одна работа, которую также стоит удостоить вниманием, – это «Комбинированная модель нейронной сети для прогнозирования цен на сырьевые товары с помощью SSA» [15]. Ее авторы изучают динамику цен на сырьевые продукты (кукурузу, золото и нефть). Одной из особенностей данной работы является то, что в ней используется сингулярно-спектральный анализ для отделения временного ряда от случайного шума, что также является одним из ключевых моментов данной курсовой работы. Для предсказания выбраны несколько различных моделей нейронных сетей. Результаты исследования свидетельствуют в пользу того, что SSA-сглаживание улучшает прогнозирование для некоторых отдельных моделей.

⁴ ETS (Error, Trend, Seasonal) – модель экспоненциального сглаживания

⁵ TBATS (Trigonometric, Box-Cox transformation, ARMA, Trend, Seasonality) – модели, основанные на экспоненциальном сглаживании

⁶ ARFIMA (autoregressive fractionally integrated moving average) – авторегрессионная дробно-интегрированная модель скользящего среднего

⁷ AR (autoregressive model) - авторегрессионная модель

⁸ VAR (vector autoregression) – модель векторной авторегрессии

⁹ BAR (Bayesian autoregression) – байесовская авторегрессия

¹⁰ BVAR (Bayesian vector autoregression) – байесовская векторная авторегрессия

¹¹ Random Walk – модель случайного блуждания

1.2. Описание метода «Гусеница»-SSA

В основе анализа валютного курса в проведенном исследовании лежит метод сингулярно-спектрального анализа. Опишем его идею и сам алгоритм.

Одной из фундаментальных работ, посвященных сингулярно-спектральному анализу, является учебник Н. Э. Голяндиной [2].

Сингулярно-спектральный анализ (Singular Spectrum Analysis, или SSA, или «Гусеница»-SSA) – это один из методов, применяемых для анализа временных рядов. Его суть заключается в разложении ряда на несколько составляющих, в результате чего задача анализа ряда трансформируется в задачу анализа его компонент. SSA позволяет выделять у ряда тренд, отыскивать периодики, сглаживание ряда, а также осуществлять полное разложение ряда на составляющие – тренд, периодики и шум.

Важным преимуществом метода является тот факт, что для его применения неважна стационарность ряда, модель тренда, периодичность переменных. Несмотря на это, метод «Гусеницы»-SSA, конечно, тоже не обделен рядом своих проблемных сторон.

Перейдем к описанию шагов данного метода.

Базовый алгоритм SSA включает в себя два этапа: разложение временного ряда и его восстановление.

Этап разложения ряда состоит из двух шагов. Первый шаг – это вложение. В результате него ряд $F = (f_0, \dots, f_{N-1})$ трансформируется в набор многомерных векторов. За L обозначается параметр, который отвечает за длину окна ($1 < L < N$), тогда после процедуры вложения формируются векторы вложения вида $X_i = (f_{i-1}, \dots, f_{i+L-2})^T, 1 \leq i \leq K$ размерности L в количестве $K = N - L + 1$ штук. В результате, получаем так называемую траекторную матрицу, столбцы которой – это полученные ранее векторы вложения:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{pmatrix} \quad (1)$$

При этом элементы матрицы \mathbf{X} на всех диагоналях, перпендикулярных главной, равны, а значит, \mathbf{X} по определению является ганкелевой матрицей¹².

Второй шаг алгоритма – *сингулярное разложение* (Singular Value Decomposition, или SVD). Пусть $\lambda_1, \dots, \lambda_L$ ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$) – это собственные значения матрицы $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, отсортированные по неубыванию, а U_1, \dots, U_L – ортонормированные собственные векторы данной матрицы, соответствующие собственным числам $\lambda_1, \dots, \lambda_L$ соответственно.

Пусть $V_i = \frac{\mathbf{X}^T U_i}{\sqrt{\lambda_i}}$, $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$, $i = 1, \dots, d$, где d – количество ненулевых собственных значений, тогда матрицы \mathbf{X}_i образуют искомое сингулярное разложение матрицы \mathbf{X} . Иными словами, оно выглядит следующим образом:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d \quad (2)$$

При этом каждая из этих матриц-составляющих является элементарными, так как их ранг равен 1.

Каждую тройку $(\sqrt{\lambda_i}, U_i, V_i)$ называют i -й собственной тройкой сингулярного разложения.

Перейдем ко второму этапу метода SSA – восстановлению. Первый шаг этого этапа (или третий шаг всего алгоритма) заключается в *группировке* всех индексов $i = 1, \dots, d$ на m непересекающихся подмножеств I_1, \dots, I_m . То есть для группы I матрица будет выглядеть так:

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p} \quad (3)$$

Выбор непересекающихся подмножеств I_1, \dots, I_m – это группировка собственных троек.

¹² Ганкелева матрица – матрица, у которой на всех диагоналях, которые перпендикулярны главной, находятся одинаковые элементы

Значит, разложение в сгруппированном виде можно сформулировать следующим образом:

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_p} \quad (4)$$

Наконец, перейдем к последнему, четвертому, шагу алгоритма. На данном этапе происходит формирование путем диагонального усреднения нового временного ряда прежней длины из полученных на предыдущем шаге матриц \mathbf{X}_{I_k} .

Опишем кратко процедуру диагонального усреднения. Обозначим за $\mathbf{Y} = (y_{ij})_{i,j=1}^{L,K}$ матрицу размера $L \times K$, за L^* и K^* – минимум и максимум из L и K соответственно. Пусть $y_{ij}^* = y_{ij}$, если $L < K$, и $y_{ij}^* = y_{ji}$ в противном случае.

Перейдем к сути диагонального усреднения. Идея состоит в том, что осуществляется перевод данной матрицы \mathbf{Y} в ряд g_0, \dots, g_{N-1} согласно правилу, по которому происходит расчет усредненных элементов вдоль диагоналей:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^*, & \text{если } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^*, & \text{если } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2}^*, & \text{если } K^* \leq k < N. \end{cases} \quad (5)$$

Диагональное усреднение применяется к матрицам \mathbf{X}_{I_k} , получаются ряды $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$, и изначальный временной ряд $F = (f_0, \dots, f_{N-1})$ переводится в сумму из m рядов:

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)}. \quad (6)$$

Итак, базовый алгоритм сингулярно-спектрального анализа рассмотрен.

Стоит также сказать о методах осуществления группировки компонент ряда. Так, например, для задачи выделения тренда или сглаживания ряда группировка может производиться в результате анализа собственных векторов: необходимо группировать те из них, которые меняются медленно. Если же цель – обнаружить колебания, то алгоритм действий основан на анализе собственных троек и понятии сильной разделимости ряда:

предполагается, что компонента F_1 является приближенно сильно отделимой от компоненты $F_2 = F - F_1$, а значит, две собственные тройки разложения соответствуют F_1 .

Если необходимо найти полное разложение ряда, то решение такой задачи может быть переформулировано в виде нахождения в отдельности тренда, колебаний, шума.

Существуют также некоторые другие способы, используемые при группировке компонент. К примеру, можно анализировать сингулярные числа, а именно, выявлять на графике собственных значений участки-плато или такие участки, которые являются менее пологими, чем соседние, и формировать тем самым группы компонент.

Наконец, следует сказать и о группировке по взвешенной корреляции. Логика метода состоит в том, что собственные тройки, образующие одну группу, зачастую соответствуют тем членам ряда, которые сильно коррелированы, и наоборот – компоненты из разных групп имеют низкую корреляцию. Это следует из понятия о приближенной разделимости двух рядов: условием данного явления служат близкие к нулю взвешенные корреляции компонент, получившихся в результате восстановления исходного временного ряда.

Глава 2. Методология курсовой работы

В данном разделе описаны методы, которые были использованы в ходе исследования.

Во-первых, для анализа временного ряда и очистки данных от шума был использован *сингулярно-спектральный анализ*, описанный в предыдущей главе.

С помощью данного метода исходный временной ряд был разложен на элементарные компоненты, которые далее были сгруппированы двумя методами: сначала это было сделано в результате *визуального анализа* составляющих ряда и поля взвешенной корреляции, а затем – более точным и строгим методом – *с помощью матрицы взвешенных корреляций*.

Также группировка была проанализирована с использованием *графика собственных чисел*.

Получаемые сгруппированные элементы в обоих случаях описанной группировки проверялись на нормальность с помощью *теста Колмогорова-Смирнова*, что позволяло сделать вывод о том, относится ли очередная полученная восстановленная компонента к шуму.

Рассмотрим данный тест более подробно. С помощью критерия Колмогорова-Смирнова осуществляется проверка гипотезы о том, что данные в выборке подчиняются некоторому закону распределения [17]. Пусть $X_n = (X_1, \dots, X_n)$ – данная выборка, которая состоит из независимых одинаково распределенных случайных величин (i.i.d.) и имеет функцию распределения $F_n(x)$. Нулевая гипотеза состоит в том, что выборка принадлежит некоторому конкретному распределению $F(x)$, альтернативная гипотеза заключается в том, что выборка не принадлежит этому распределению. Если функция распределения $F(x)$ непрерывна, то можно применить тест Колмогорова-Смирнова [8]. Значение статистики рассчитывается по следующей формуле:

$$D_n = \sup_x |F_n(x) - F(x)|. \quad (7)$$

Наконец, после получения очищенного от шума ряда была применена *интегрированная модель авторегрессии скользящего среднего*.

Модель ARMA является объединением моделей авторегрессии (AR) и модели скользящего среднего (MA) [13]. При этом модель ARMA может быть использована только для стационарных рядов. Поэтому Бокс и Дженкинс создали модель ARIMA, которая делает возможным работать в том числе и с некоторыми нестационарными рядами, а именно, с теми, которые в результате взятия d последовательных разностей приводятся к стационарным рядам [6], для которых уже применима модель ARMA.

Несезонные модели ARIMA (то есть такие, в которых не рассматривается возможное наличие сезонности в данных) имеют три параметра: p – порядок авторегрессии, d – порядок разности, q – порядок скользящего среднего [3].

Модель ARIMA можно записать в следующем виде [1]:

$$\Delta^d y_t = c + \sum_{i=1}^p a_i \Delta^d y_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t, \quad (8)$$

где Δ^d – оператор разности, c, a, b – параметры модели, ε_t – стационарный временной ряд.

По графику временного ряда обычно представляется невозможным понять, какими должны быть значения параметров модели. Иногда в определении значений параметров p и q могут помочь *графики ACF¹³ и PACF¹⁴* [13].

При этом используются следующие свойства ACF и PACF при подборе наиболее оптимальных параметров:

Таблица 1

Свойства ACF и PACF для определения параметров модели

ACF	PACF	Модель
Экспоненциально затухает	Выброс только на первом лаге	AR (1) (модель авторегрессии 1-го порядка)

¹³ ACF (autocorrelation function) – АКФ, автокорреляционная функция

¹⁴ PACF (partial autocorrelation function) – ЧАКФ, частичная автокорреляционная функция

Экспоненциально затухает либо имеет форму затухающей синусоидальной волны	Выброс только на 1 и 2 лагах	AR(2) (модель авторегрессии 2-го порядка)
--	---------------------------------	---

Продолжение таблицы 1

Имеется выброс на первом лаге	Экспоненциально затухает	МА (1) (модель скользящего среднего первого порядка)
Имеются выбросы на первом и втором лагах	Экспоненциально затухает либо имеет форму синусоидальной волны	МА (2) (модель скользящего среднего первого порядка)
Экспоненциально затухает с первого лага	Экспоненциально затухает с первого лага	ARMA (1,1) (авторегрессионная модель скользящего среднего первого порядка)

Источник: Красулин А. А. Имитационное моделирование динамических систем со случайными компонентами //Имитационное моделирование. Теория и практика (ИММОД 2013). – 2013. – С. 313-317.

В работе была применена модель AutoARIMA¹⁵, которая не требует знания точных значений параметров в отличие от просто ARIMA, в которой требуется задать четкие значения для параметров модели.

Возвращаясь к временному ряду валютной пары, AutoARIMA была применена к исходному неочищенному ряду, затем к ряду, полученному после визуальной группировки компонент и, наконец, к ряду, который получился после группировки по корреляции.

¹⁵ AutoARIMA – автоматическая модель ARIMA

Для каждого из трех случаев были вычислены *метрики качества*: $RMSE^{16}$, MAE^{17} , $MAPE^{18}$ и сделан вывод о лучшей модели и, соответственно, о том, верно ли было сделанное предположение, что на очищенных от шума данных модель будет работать лучше.

$RMSE$, MAE и $MAPE$ являются наиболее распространенными метриками качества ввиду простоты их вычисления и понятности, а также независимости от свойств ошибок [7].

Первые две из них являются относительными метриками качества (то есть полученные значения данных ошибок зависят от единиц измерения исходных данных). Следовательно, эти два показателя используются при сравнении прогнозных значений, полученных различными методами на одних и тех же данных или при сопоставлении качества одной и той же модели на нескольких массивах данных, измеряемых в единой шкале [13]. Чем меньше корень из средней квадратичной ошибки и средней абсолютной ошибки, тем лучше прогноз [7].

Итак, $RMSE$ основана на вычислении среднеквадратической ошибки, MAE – абсолютных значений ошибок [13]:

$$MAE = mean(|e_t|) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (9)$$

$$RMSE = \sqrt{mean(e_t^2)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (10)$$

$MAPE$ вычисляет абсолютное значение ошибки (выражается в процентах или доле):

$$MAPE = 100\% \cdot \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} \quad (11)$$

Это дает возможность сравнивать модели с различными масштабами измерения. Наименьшее значение данной ошибки также указывает на лучшее качество модели.

¹⁶ $RMSE$ (Root Mean Squared Error) – корень из средней квадратичной ошибки

¹⁷ MAE (Mean Absolute Error) – средняя абсолютная ошибка

¹⁸ $MAPE$ (Mean Absolute Percent Error) – абсолютная процентная ошибка

Описанный выше алгоритм действий при выполнении работы был реализован с помощью языка программирования Python. Опишем подробнее, какие основные инструменты использовались при написании кода.

Во-первых, для работы с данными использовались библиотеки анализа данных `numpy` и `pandas`.

Для визуализации котировок в виде графика свечей использовалась библиотека `plotly` [22], с помощью которой можно довольно легко создавать различные интерактивные графики. В результате, получилось отобразить динамику изменения исследуемой валютной пары с возможностью изменять период отображаемых данных, не меняя код.

Библиотека `matplotlib` [21] также использовалась в ходе работы: она пригодилась при отображении графиков компонент временного ряда, а также при создании так называемого поля корреляции (в данном случае использовались взвешенные корреляции), которое послужило фундаментом для визуальной группировки компонент ряда.

Для построения модели Auto-ARIMA была использована библиотека `sktime` [23]. Авто-АРИМА проводит различные тесты на определение параметра порядка разности d (такие как тест Дики-Фуллера, Филлипса-Перрона, Квятковского-Филлипса-Шмидта-Шина), а затем подгоняет модель под заданный диапазон параметров. Чтобы найти лучшую модель из возможных для конкретного временного ряда Auto-ARIMA использует различные информационные критерии: AIC^{19} , $SAIC^{20}$, BIC^{21} , а также $HQIC^{22}$. Используется модель с наименьшим значением критериев.

¹⁹ AIC (Akaike information criterion) – информационный критерий Акаике

²⁰ SAIC (Akaike's information criterion corrected) – скорректированный информационный критерий Акаике

²¹ BIC (Bayesian information criterion) – Байесовский информационный критерий

²² HQIC (Hannan–Quinn information criterion) – информационный критерий Ханнана-Куинна

Глава 3. Применение сингулярно-спектрального анализа в задаче отделения ряда от шума и прогнозирования

3.1. Описание выбранных данных

Для анализа и прогнозирования валютного курса евро/доллар были взяты подневные данные по цене открытия, цене закрытия, максимальной и минимальной ценам, а также по объему торгов валютной пары за 3 месяца: с 1 февраля по 30 апреля 2021 года (то есть данные за каждый день, всего получилось 77 значений – из-за наличия пропусков в данных в выходные дни). Данные брались из раздела «Теханализ» с сайта finam.ru [18].

Остановимся чуть подробнее на некоторых из этих финансовых показателей активов.

Цена открытия представляет собой стоимость ценной бумаги в начале торгов. Поэтому обычно она считается той ценой, на которую еще не успели повлиять внешние факторы [20]. Поэтому данный показатель составляет основу анализа выбранной валютной пары в курсовой работе. Цена открытия используется для составления временного ряда и его анализа, разложения и группировки ряда, прогнозирования валютного курса.

Цена закрытия отражает цену последней сделки, совершенной непосредственно перед закрытием торгов на бирже [20]. Таким образом, можно сказать, что по данному показателю можно сделать некоторые выводы об ожиданиях агентов на фондовом рынке. При этом цена открытия «завтра» может отличаться от цены закрытия «сегодня», так как котировки подвержены изменениям, даже когда биржа закрыта.

Данный показатель вместе с другими используется на этапе первичного анализа массива данных, для построения так называемого графика японских свечей – графика, отражающего изменения биржевых котировок.

Максимальная и минимальная цена также являются важными биржевыми показателями, они характеризуют диапазон цен, в котором происходили торги в течение торгового дня. Иными словами, данный показатель – это есть

максимум и минимум той цены, которая была уплачена за акцию в течение торгового дня на бирже.

После проведения некоторой предобработки данных для дальнейшей работы, стало возможным осуществить визуализацию. Ниже приведен график японских свечей, которые отражают изменения котировок валютного курса доллар/рубль, сделано это было с помощью питоновской библиотеки plotly (код работы представлен на интернет-ресурсе github [16]). Написанный код строит интерактивный график, так что имеется возможность выбрать определенный период для отображения.



Рисунок 1. Интерактивный котировок валютного курса евро/доллар

Источник: построение автором

3.2. Применение сингулярно-спектрального анализа и группировка компонент

Перейдем к применению сингулярно-спектрального анализа. Для осуществления разложения ряда на компоненты за основу был взят код с Kaggle [19] с некоторыми изменениями, автор которого опирался на книгу [2] Н. Я. Голяндиной, которая является одной из фундаментальных работ по теме. Функция, написанная автором кода, осуществляет разложение с помощью метода SSA, вывод элементарных компонент ряда, расчет и визуализацию матрицы взвешенной корреляции. На вход принимается непосредственно сам временной ряд и длина окна L .

Для последующего анализа была взята цена открытия валютной пары. Ряд был разложен на 3, 26 ($\approx N/3$) и 38 ($\approx N/2$) компонент.

На рисунках 2а) и 2б) представлены график сгруппированных компонент в случае разложения ряда на 3 составляющие и взвешенная корреляция между компонентами.

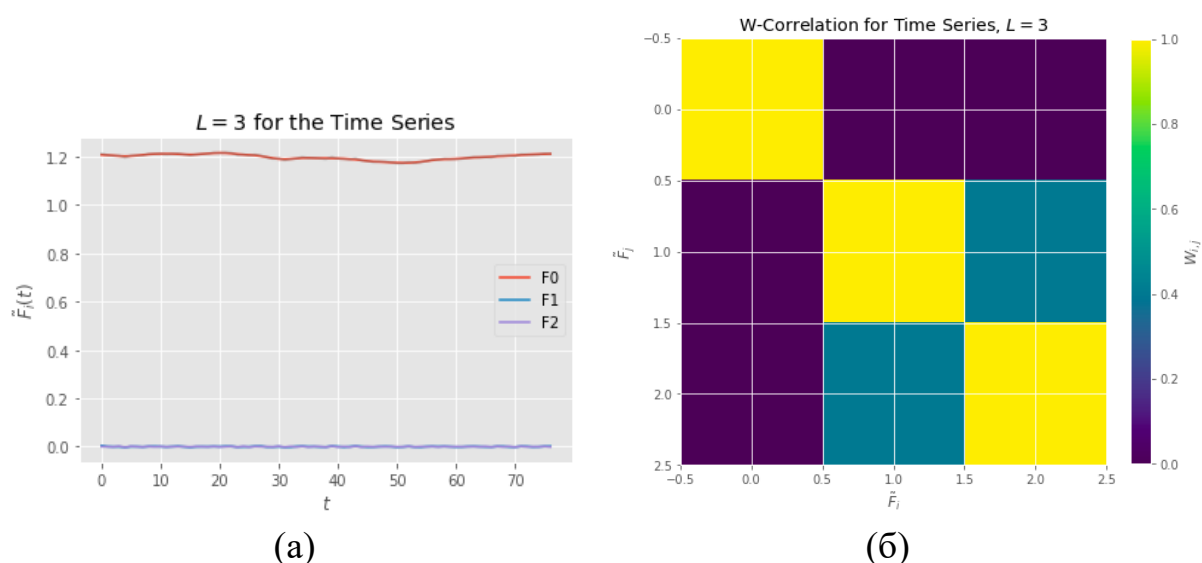


Рисунок 2. График компонент ряда при $L = 3$ (а) и корреляционное поле (б)

Источник: построение автором

Аналогичные графики были построены при разложении на 26 и 28 компонент.

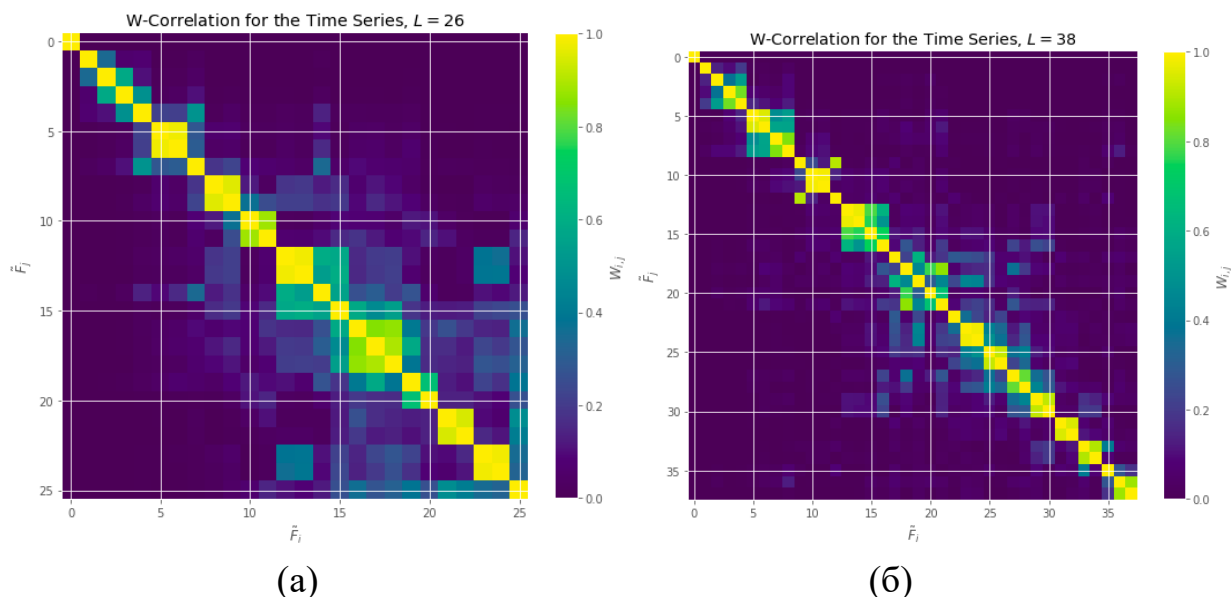


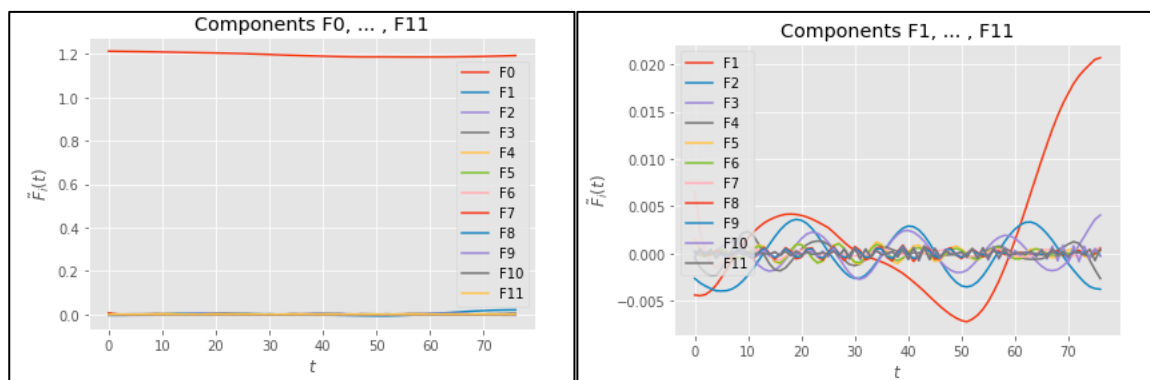
Рисунок 3. Корреляционное поле при $L = 26$ (а) и при $L = 38$ (б)

Источник: построение автором

Визуализация поля корреляции дает возможность увидеть и легко определить связь между компонентами, а значит, она дает возможность предположить возможную группировку выделенных составляющих ряда.

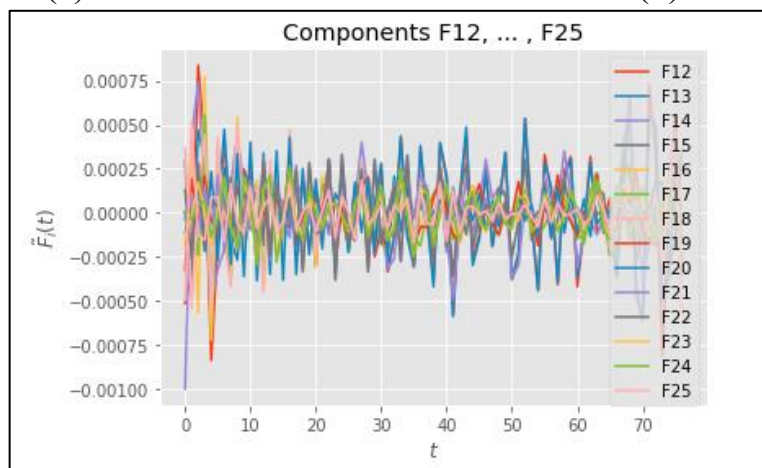
Например, при окне 26 видно, что с 12 по 25 компоненту наблюдается некоторая скоррелированная область, которую можно было бы объединить в одну компоненту и предположить, что она будет являться шумом [2], а все оставшиеся – в другую.

Можно взглянуть на графики отдельно для компонент с 0-ой по 11-ую и с 12-ой по 25-ую и заметить, что первая группа компонент имеет по большей части все же более плавную динамику (представлена на рисунках 4(а) и 4(б), второй график – без нулевой компоненты, для близкого рассмотрения), нежели вторая группа, которая в свою очередь, как видно на рисунке 4(в), имеет выраженные сильные колебания, что может говорить о том, что вторая компонента скорее всего является шумом).



(a)

(б)



(в)

Рисунок 4. Графики компонент временного ряда с 0 по 11 (а), с 1 по 11 (б) и с 12 по 25 (в)

Источник: построение автором

Итак, на основе полученного поля взвешенной корреляции и визуального анализа компонент, можно предположить, что элементы с 0 по 11 и с 12 по 25 должны быть сгруппированы. В результате группировки получаем следующие графики восстановленных компонент:

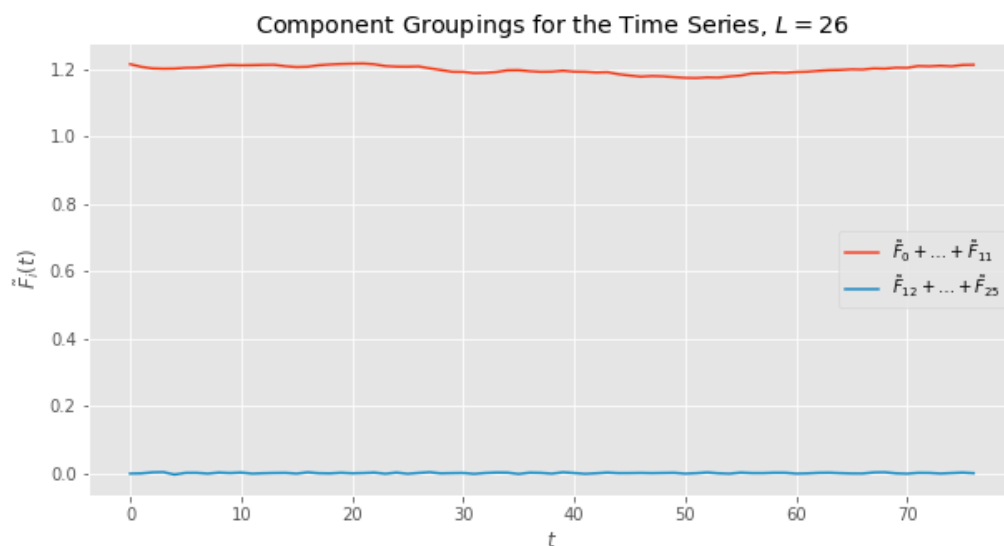


Рисунок 5. График сгруппированных после визуального анализа компонент

Источник: построение автором

Посмотрим, как соотносены между собой график исходного временного ряда и график сгруппированных с 0-ую по 11-ую компонент:

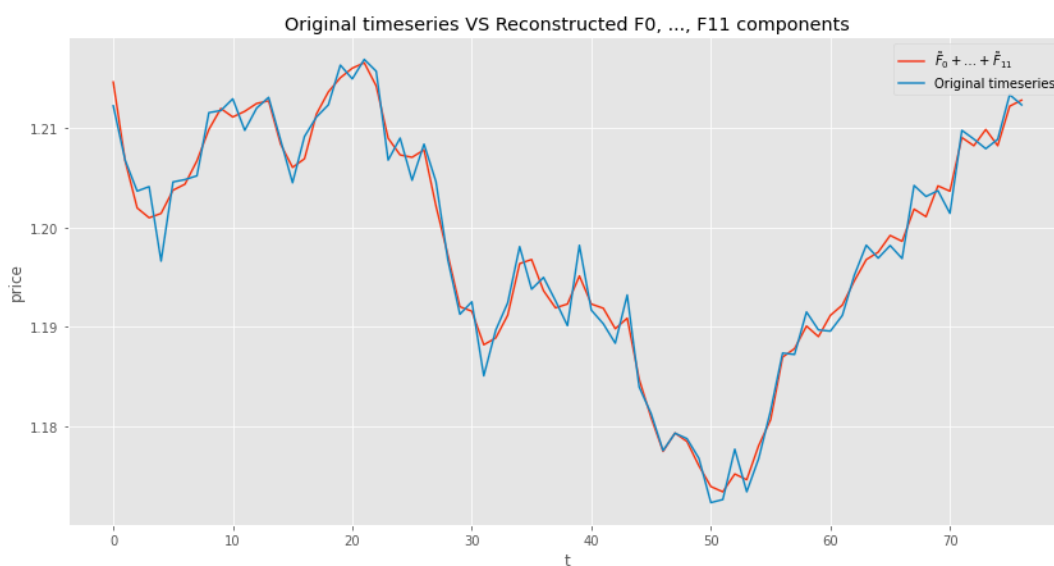


Рисунок 6. График исходного временного ряда и график сгруппированных с

0-ую по 11-ую компонент

Источник: построение автором

Полученная кривая является некоторым улучшением исходной, чего нельзя сказать о второй компоненте: ее график (см. рисунок 7) сильно отличается от исходного, поэтому можно предположить, что вторая компонента как раз-таки является шумом.

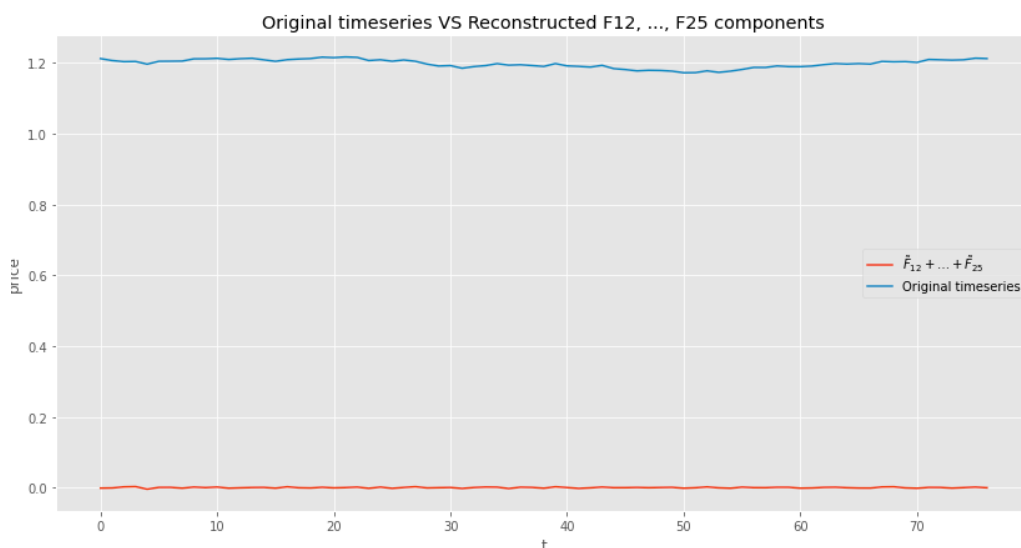


Рисунок 7. График исходного временного ряда и график сгруппированных с 12-ую по 25-ую компонент

Источник: построение автором

Чтобы убедиться в представленных выше наблюдениях, проверим группированные компоненты на нормальность с помощью теста Колмогорова-Смирнова (также называемый Lilliefors' test).

Получаем результаты, которые представлены в таблице ниже:

Таблица 2

Результаты теста Колмогорова-Смирнова, полученные в Python

	Значение статистики	p-value	p-value < 0.05?	Нулевая гипотеза
F0, ..., F11	0.1100101871018 9181	0.0350441348399 8005	да	отвергается
F12, ..., F25	0.0524843456907 71705	0.9054215933268 338	нет	не отвергается

Источник: расчеты автора

Значит, поскольку для первой компоненты нулевая гипотеза о нормальном распределении отвергается с вероятностью ошибки 0,05, а для второй – не отвергается на том же уровне значимости, то наши предположения о том, что первая компонента отражает тренд и периодики, а вторая – шум, оказались верными.

Несмотря на полученные выводы, визуальный способ группировки, безусловно, воспринимается скорее всего как довольно неточный и нематематичный метод. Поэтому стоит рассмотреть другой способ отделения ряда от шума.

Будем группировать компоненты, опираясь на матрицу взвешенных корреляций между этими компонентами. Для этого была написана функция, которая проверяла в поданном на вход датафрейме наличие компонент, у которых попарная корреляция была, для начала, выше порогового значения 0,9. В случае выявления таких компонент, производилась их группировка и перерасчет матрицы взвешенных корреляций уже для полученного нового датафрейма, в котором были сгруппированы эти компоненты. Далее для нового датафрейма и соответствующей ему новой матрицы снова происходила проверка на наличие компонент с корреляцией выше изначального порога 0,9.

Если же таких компонент не было найдено, то в матрице искались значения выше следующего порогового значения – 0,8, далее осуществлялись аналогичные действия (как с первым порогом). И, так же как с прежним порогом, при отсутствии таких компонент происходил переход к следующему порогу – 0,7, а от него – к 0,6 и, наконец, к 0,5, в отсутствии значений выше которого функция прекращала свою работу.

В результате, из исходных 26 компонент осталась лишь половина – 13: F0, F1, F25, F24 + F23, F6 + F5, F8 + F9, F22 + F21, F10 + F11, F18 + F17 + F16, F19 + F20, F2 + F3, F7 + F4, F13 + F12 + F15 + F14.

Ранее было упомянуто, что один из возможных методов группировки компонент временного ряда основан на анализе собственных значений. Выявляя участки, на которых наблюдается плато, можно сделать вывод о необходимости группировки тех или иных компонент. Посмотрим, каким образом результаты, полученные группировкой по корреляции, соотносятся с графиком собственных значений. Нулевая компонента получается сильно оторванной от остальных, посмотрим без нее:

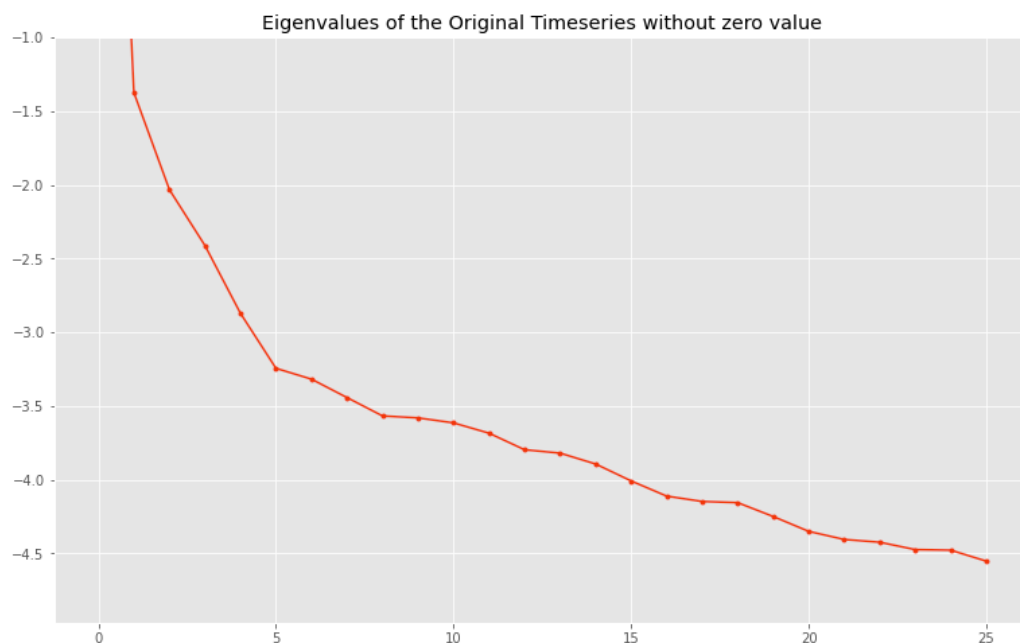


Рисунок 8. График собственных значений в результате разложения временного ряда без нулевого значения

Источник: построение автором

Во-первых, сильный отрыв нулевой компоненты от остальных – признак того, что скорее всего это будет отдельная компонента. То же самое можно сказать про компоненту F1. Далее, компоненты F2 и F3, объединенные ранее группировкой по корреляции: в целом, если сравнивать наклон графика до второй и после третьей компоненты, то между F2 и F3 график является более пологим, значит, их группировку можно обосновать. Следующей сгруппированной парой была F4 + F7, в данном случае график собственных значений не может объяснить логику такого объединения. Группировка 5 и 6 компонент вполне логична: график на данном участке более пологий чем до и после него. Далее по корреляции были попарно объединены 8 и 9, 10 и 11 компоненты, однако, по графику можно было бы предположить, что все они относятся к одной компоненте, а не к двум разным. Также это выглядит нелогичным, учитывая, что следующая группа по корреляции уже включает 4 компоненты: 12, 13, 14 и 15, хотя эта группа компонент на графике выглядит по форме очень похоже на предыдущие четыре, которые не объединены в одну большую группу. Тем не менее, есть и однозначно логичные объединения:

группа с 16 по 18 компоненту на графике имеет выраженную одинаковую пологость, как и в случае с парами 21 и 22, 23 и 24.

Вернемся к группировке по взвешенной корреляции. Посмотрим на графики полученных компонент:

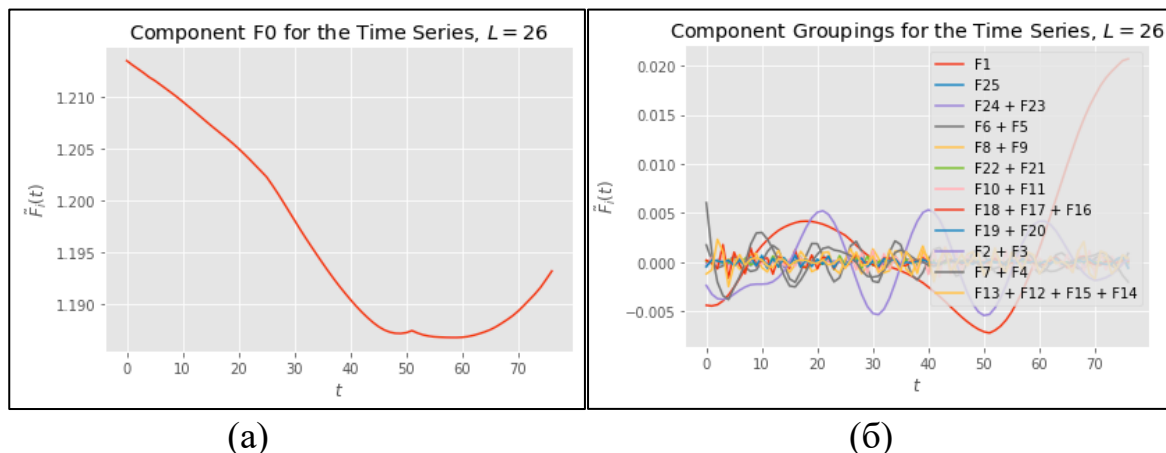


Рисунок 9. Графики нулевой компоненты (а) и всех остальных (б)

Источник: построение автором

На рисунке 9(а) представлен график нулевой компоненты, на рисунке 9(б) – всех остальных после группировки, так как ее исключение позволяет лучше разглядеть другие графики.

Как видно, есть компоненты, которые имеют довольно плавную динамику, а есть те, которые похожи на колебания. Последние скорее всего относятся к шуму.

Проанализируем каждую из восстановленных компонент ряда на нормальность с помощью критерия Лиллиефорса, результаты расчета которого представлены в таблице ниже:

Таблица 3

Результаты тестирования остатков на нормальность

Восстановленная компонента	Значение статистики	p-value	Нулевая гипотеза
F0	0,17375650697174405	0,00099999999999998899	Отвергается
F1	0,19026461379100512	0,00099999999999999899	Отвергается
F25	0,10772275006824818	0,04231201708558177	Отвергается

F24 + F23	0,06816331401820 663	0,56996630596050 32	НЕ отвергается
F6 + F5	0,07526707277096 695	0,41661083910765 173	НЕ отвергается
F8 + F9	0,06154641765556 4956	0,72853734866988 43	НЕ отвергается
F22 + F21	0,11920218262783 933	0.01554501919487 4799	Отвергается
F10 + F11	0,06178609884329 611	0.72279349327705 49	НЕ отвергается
F18 + F17	0,08231465614084 266	0.27615528768522 735	НЕ отвергается
F19 + F20	0,06622962184491 388	0.61630639765805 34	НЕ отвергается
F2 + F3	0,11151628906987 926	0.03025879142190 756	Отвергается
F7 + F4	0,09214317403360 062	0.15243251034993 485	НЕ отвергается
F13 + F12 + F15 + F14	0,05386618453626 968	0.88154694499373 98	НЕ отвергается

Источник: расчеты автора

При значении p -уровня значимости меньше 0,05, гипотеза о принадлежности данных нормальному распределению отвергалась. В результате, получилось 5 компонент, которые не подчиняются нормальному распределению (это компоненты F0, F1, F25, F22 + F21, F2 + F3) и 8 нормально распределенных составляющих.

Теперь можно посмотреть на графики компонент в отдельности по тем, которые нормально распределены и тем, которые не подчиняются нормальному закону (компонента F0 не представлена на графике, так как она вновь не дает увидеть динамику других компонент):

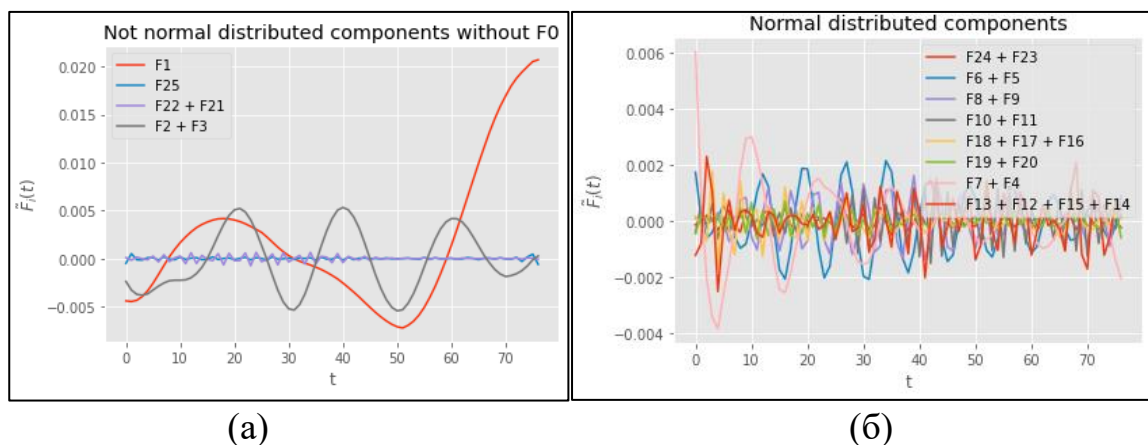


Рисунок 10. Графики компонент, не подчиняющихся нормальному закону (а) и нормально распределенных (б)

Источник: построение автором

На графике справа все компоненты представляют собой случайные колебания, на левом же компоненты не похожи на шум, за исключением, может быть, компонент F25 и F22 + F21, что неудивительно: в представленной ранее таблице для них отмечались значения уровня значимости наиболее близкие к граничному p-value.

Далее, компоненты с нормальным распределением нужно отбросить от временного ряда и рассматривать временной ряд из группировки оставшихся компонент.

Проверка полученного ряда на нормальность дает также отвержение гипотезы о нормальности.

В завершение, посмотрим на соотношение графиков исходного временного ряда и ряда после удаления шума (см. рисунок 11). В целом, можно увидеть, что график очищенного временного ряда отображает динамику исходного временного ряда, сглаживая имеющиеся колебания.

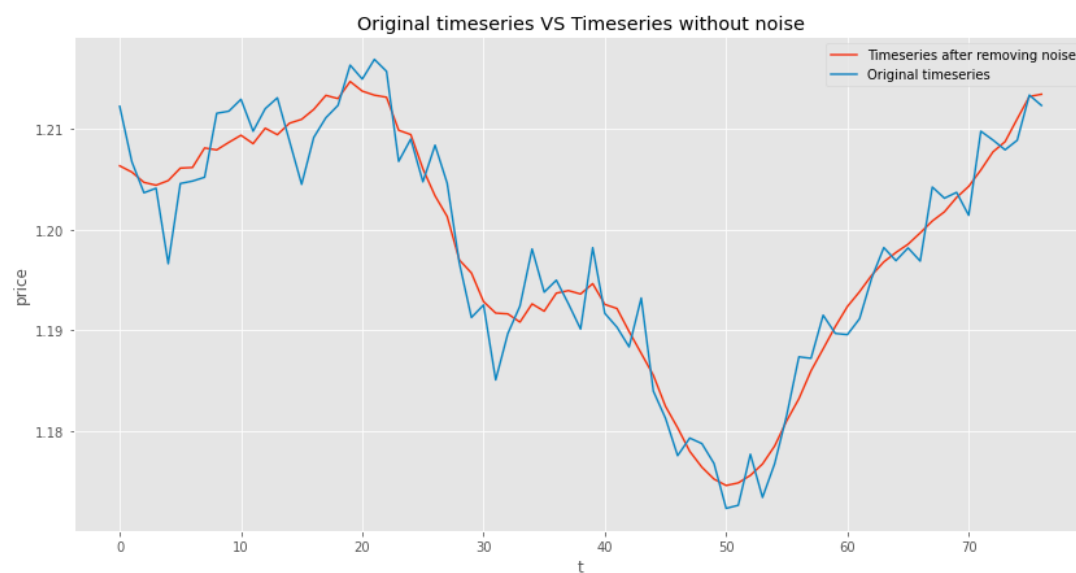


Рисунок 11. Исходный временной ряд и ряд, очищенный после группировки по корреляции

Источник: построение автором

3.3. Применение модели ARIMA для прогнозирования динамики валютного курса

Последним этапом работы было применение модели прогнозирования AutoARIMA к исходному временному ряду, к ряду после удаления шума в результате визуального выявления групп компонент и, наконец, к ряду, очищенному от шума после группировки компонент посредством выявления больших значений попарных взвешенных корреляций.

Каждая из выборок была разделена на обучающую и тестовую выборку, которые были взяты в количестве 90% и 10% соответственно. Такие объемы выборок были взяты исходя из того предположения, что трейдеры, участвующие в торгах на финансовом рынке вряд ли захотят спланировать свои действия на рынке на такой большой период, как месяц или даже полмесяца. Кроме того, валютный курс является довольно волатильной величиной, поэтому объем тестовой выборки взят в малом объеме (а не, например, 20% или 30%, которые обычно используются).

Итак, модель была обучена на соответствующих обучающих выборках.

Посмотрим сначала на результаты для исходного временного ряда.

Таблица 4

Отчет об обучении модели AutoARIMA на исходных данных

Dep. Variable:		y		No. Observations:		69
Model:		SARIMAX(0, 1, 0)		Log Likelihood		276,496
Sample:		0		AIC		-548,993
		-69		BIC		-544,554
Covariance Type:		opg		HQIC		-547,234
	coef	std err	z	P> z 	[0,025	0,975]
Intercept	-0,0001	0,001	-0,265	0,791	-0,001	0,001
sigma2	1,721e-05	3,51e-06	4,905	0,000	1,03e-05	2,41e-05
Ljung-Box (L1) (Q):		1,33		Jarque-Bera (JB):		1,13

Prob(Q):	0,25	Prob(JB):	0,57
Heteroskedasticity (H):	0,70	Skew:	-0,14
Prob(H) (two-sided):	0,40	Kurtosis:	2,43

Источник: расчеты автора

Как видно из представленных о модели данных, AutoARIMA использовала значения параметров $p = 0$, $d = 1$, $q = 0$. Модель с таким параметра является моделью случайного блуждания.

Проанализируем графики автокорреляционной функции и частичной автокорреляционной функции, представленные на рисунке 12. Так как на графике автокорреляции пик происходит при 1 лаге, а на графике частичной автокорреляционной функции можно увидеть некоторое подобие затухающей синусоидальной волны, то оптимальной будет модель с параметрами $p = 0$, $q = 1$. Однако, будем доверять полученной модели, которая более строго осуществляет поиск оптимальных значений параметров. К тому же, второй график не имеет четкой тенденции к затуханию, так как на нем все же наблюдаются некоторые внезапные выбросы значений.

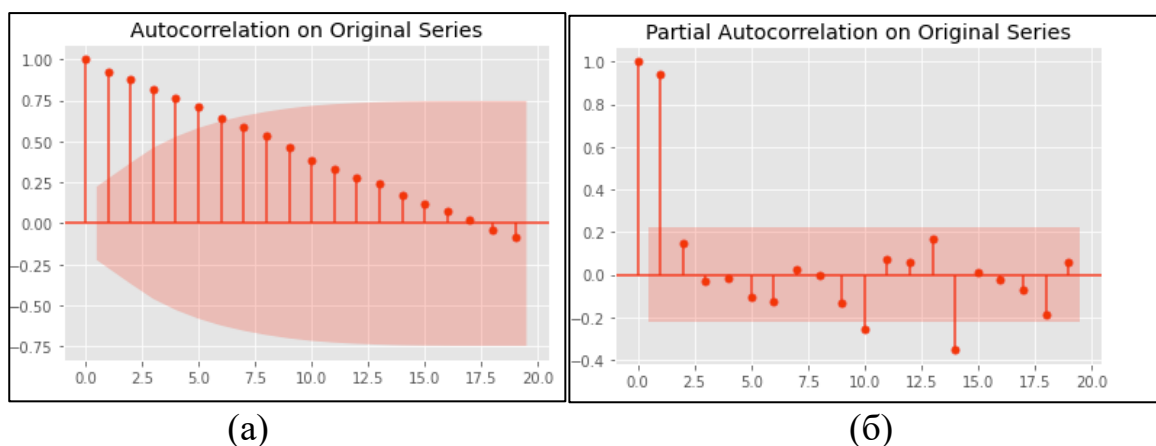


Рисунок 12. График автокорреляционной функции (а) и частичной автокорреляционной функции (б) для исходного временного ряда

Источник: построение автором

В отчете, полученном после обучения модели, представлены оценки коэффициентов модели, их значимость, а также различные тесты. Например, здесь есть значения так называемых информационных критериев: критерий

Акаике (AIC) и критерий Шварца (BIC). Суть их расчета заключается в том, что чем меньше значение данных критериев, тем лучше модель. С помощью них как раз-таки осуществляется поиск лучшей модели, то есть определение оптимального значения параметров p , d , q модели с учетом, с одной стороны, получаемой точности модели, а с другой – ее сложности. Сами значения данных критериев не являются информативными, а рассматриваются только для сравнения различных моделей, обучаемых на некоторой выборке [11], поэтому в данном случае данные показатели не особо интересны.

Также можно обратить внимание на тестирование остатков. Во-первых, на Q-тест Льюнга-Бокса, который применяется для проверки случайности остатков, то есть их принадлежности «белому шуму» [5]. Нулевая гипотеза состоит в том, что данные случайны. Так как в данном случае получилось $p\text{-value} = 0,25$, что больше уровня значимости, например, $0,05$, то гипотеза о случайности остатков не отвергается, то есть автокорреляция остатков отсутствует. Во-вторых, в отчете есть также тест Жарка-Бера, который проверяет остатки на нормальность. Нулевая гипотеза о нормальности остатков в данном случае не отвергается (так как $p\text{-value} = 0,57 > 0,05$).

Не сравнивая с моделями, обученными на очищенных данных, сложно сделать какие-то выводы о качестве одной модели по метрикам. Поэтому проанализируем значения RMSE, MAE и MAPE в конце.

Посмотрим теперь на сводку по обучению модели на данных после визуальной группировки компонент ряда (см. таблицу 5).

Таблица 5

Отчет об обучении модели AutoARIMA после очистки данных визуальной группировкой компонент

Dep. Variable:	y	No. Observations:	69
Model:	SARIMAX(1, 1, 2)	Log Likelihood	324,693
Sample:	0	AIC	-639,386
	-69	BIC	-628,288
Covariance Type:	opg	HQIC	-634,989

	coef	std err	z	P> z 	[0,025	0,975]
Intercept	-0,0005	0,001	-0,586	0,558	-0,002	0,001
ar.L1	-0,7167	0,131	-5,462	0,000	-0,974	-0,460
ma.L1	1,2734	0,157	8,137	0,000	0,967	1,580
ma.L2	0,8103	0,137	5,901	0,000	0,541	1,079
sigma2	4,015e-06	8,4e-07	4,779	0,000	2,37e-06	5,66e-06
Ljung-Box (L1) (Q):				2,27	Jarque-Bera (JB):	5,79
Prob(Q):				0,13	Prob(JB):	0,06
Heteroskedasticity (H):				0,52	Skew:	-0,69
Prob(H) (two-sided):				0,12	Kurtosis:	3,35

Источник: расчеты автора

Итак, была выбрана модель с параметрами $p = 1$, $d = 1$, $q = 2$. Посмотрим, как это соотносится с графиками автокорреляции и частичной автокорреляции:

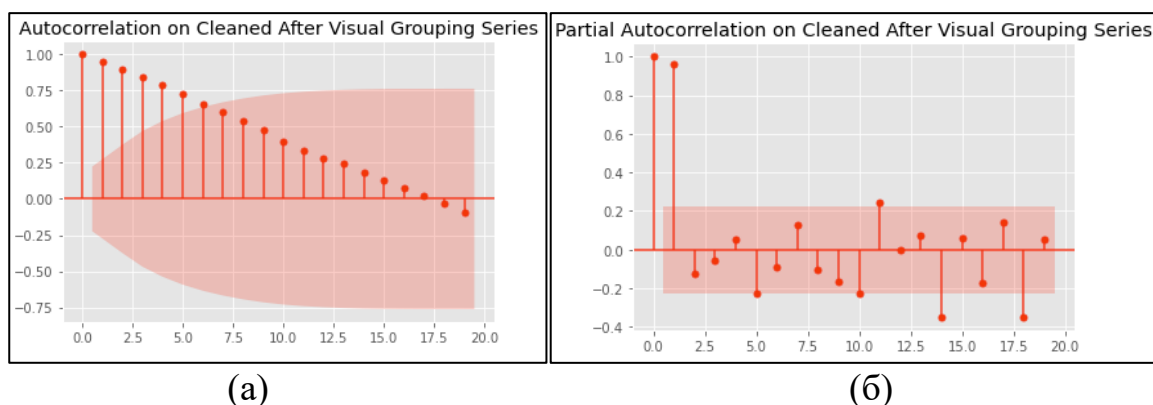


Рисунок 13. График автокорреляционной функции (а) и частичной автокорреляционной функции (б) для очищенного после визуальной группировки ряда

Источник: построение автором

В целом, выбор значения параметра модели скользящего среднего ($q = 2$) выглядит разумно: АКФ имеет выбросы на первом или первых двух лагах, ЧАКФ экспоненциально затухает в начале, а далее имеет форму

синусоидальной волны, что говорит о том, что это скорее модель авторегрессии.

Аналогичным образом, в отчете представлены оценки коэффициентов модели (теперь уже появляются коэффициенты для AR и MA частей, которых не было ранее из-за нулевых оптимальных p и q), их значимость, результаты тестирования остатков на нормальность, случайность, гетероскедастичность, а также информационные критерии, полученные для данной модели.

Как и для первого случая, по полученному p -value можно сделать выводы о том, что остатки случайны, для них отсутствует гетероскедастичность, и они подчинены нормальному распределению.

Наконец, посмотрим на модель, обученную на данных, очищенных от шума в результате группировки компонент по матрице взвешенных корреляций (см. таблицу 6).

Таблица 6

Отчет об обучении модели AutoARIMA после очистки данных группировкой компонент по матрице взвешенных корреляций

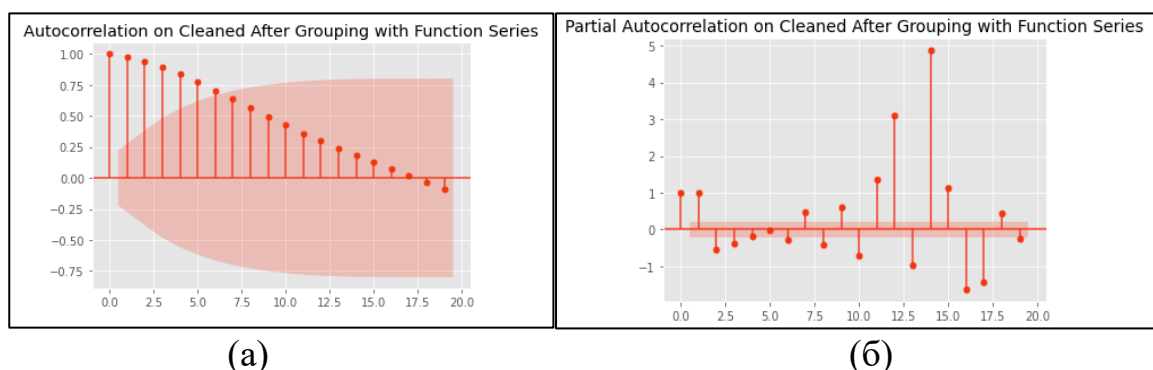
Dep. Variable:		y		No. Observations:		69
Model:		SARIMAX(4, 1, 0)		Log Likelihood		414,029
Sample:		0		AIC		-816,058
		-69		BIC		-802,741
Covariance Type:		opg		HQIC		-810,782
	coef	std err	z	P> z 	[0,025	0,975]
Intercept	-3,054e-05	7,7e-05	-0,396	0,692	-0,000	0,000
ar.L1	0,2479	0,067	3,717	0,000	0,117	0,379
ar.L2	1,1299	0,076	14,910	0,000	0,981	1,278
ar.L3	0,2524	0,079	3,189	0,001	0,097	0,408
ar.L4	-0,8381	0,069	-12,097	0,000	-0,974	-0,702
sigma2	2,819e-07	4,81e-08	5,862	0,000	1,88e-07	3,76e-07
Ljung-Box (L1) (Q):		0,01		Jarque-Bera (JB):		1,57
Prob(Q):		0,92		Prob(JB):		0,46

Heteroskedasticity (H):	0,38	Skew:	-0,31
Prob(H) (two-sided):	0,02	Kurtosis:	3,40

Источник: расчеты автора

В данном случае параметр авторегрессионной части: $p = 4$, порядок разности: $d = 1$, порядок скользящего среднего: $q = 0$.

По графикам трудно предположить значения параметров: ЧАКФ не имеет выбросов на первых двух лагах, не выглядит как экспоненциальное затухание или синусоидальная волна. По АКФ можно увидеть, что выбросы наблюдаются на первых лагах, поэтому в целом, выбор модели кажется логичным.



(a)

(б)

Рисунок 14. График автокорреляционной функции (а) и частичной автокорреляционной функции (б) для очищенного по матрице взвешенных корреляций временного ряда

Источник: построение автором

Тестирование остатков снова показало, что они являются случайными, нормально распределенными (на уровне значимости 0,05), гомоскедастичными (на уровне значимости 0,01).

Наконец, можно посмотреть и проинтерпретировать рассчитанные для трех случаев показатели качества:

Таблица 7

Показатели качества моделей, полученные в Python

	RMSE	MAE	MAPE
--	------	-----	------

Исходный временной ряд	0,007042	0,006132	0,506624
------------------------	----------	----------	----------

Продолжение таблицы 7

Временной ряд после визуальной группировки	0,009126	0,00832	0,687646
Временной ряд после группировки по матрице взвешенных корреляций	0,005628	0,004454	0,367736

Источник: расчеты автора

Для данных после очищения от шума визуальной группировкой получилось наихудшее качество модели, то есть даже хуже, чем для исходного временного ряда, что было вполне ожидаемо: визуальная группировка является довольно неточным методом.

Для модели, обученной на временном ряде после группировки с помощью функции, выявляющей максимальные корреляции в матрице взвешенных корреляций, получилось наилучшее качество: RMSE, MAE и MAPE – наименьшие.

Выводы

Итак, в ходе работы был изучен и применен к временному ряду цены открытия валютной пары евро/доллар метод сингулярно-спектрального анализа, выполнена группировка компонент временного ряда визуальным анализом поля корреляции и по матрице взвешенных корреляций, сопоставлены полученные компоненты с графиком собственных значений, и наконец, была обучена модель AutoARIMA на начальных и полученных очищенных от шума временных рядах.

В результате сравнения моделей, обученных на очищенных и исходных данных, получилось, что при аккуратной очистке (путем восстановления компонент по матрице взвешенных корреляций) показатели качества модели получаются лучше, тогда как грубая визуальная оценка возможных группировок компонент привела к ухудшению качества прогноза.

Таким образом, в целом, выдвинутая в начале работы гипотеза о том, что очистка данных от шума с помощью метода SSA приведет к улучшению качества модели, подтвердилось проведенным исследованием (если говорить о строгом методе группировки). Значит, применение участником финансового рынка сингулярно-спектрального анализа к ценам валютного курса евро/доллар должно приводить к более эффективной торговле.

Дальнейшее развитие исследования может заключаться в применении других методов группировки составляющих ряда, использовании других форматов данных (например, поминутных) и применении иных моделей прогнозирования и сравнении с уже используемой ARIMA.

Список использованной литературы

1. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление: Пер. с англ. //Мир. – 1974. – Т.1. – 406 с.
2. Голяндина Н. Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. //СПбГУ. – 2004. – 76 с.
3. Доронина А. И. Модели временного ряда: AR (p), MA (q), ARIMA (p, d, q). Пример исследования потребления нефтепродуктов во Франции //Финансовый Университет при Правительстве Российской Федерации. – 2014.
4. Красулин А. А. Имитационное моделирование динамических систем со случайными компонентами //Имитационное моделирование. Теория и практика (ИММОД 2013). – 2013. – С. 313-317.
5. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика: начальный курс. //Дело. – 1997. – 576 с.
6. Трегуб А. В., Трегуб И. В. Методика построения модели ARIMA для прогнозирования динамики временных рядов //Вестник МГУЛ – Лесной вестник. – 2011. – №. 5. – С. 179-183.
7. Турунцева М. Ю. Оценка качества прогнозов: простейшие методы //Российское предпринимательство. – 2011. – №. 8-1. – С. 50-56.
8. Чернова Н. И. Математическая статистика: Учеб. пособие //Новосибирский государственный университет. Новосибирск. – 2007. – 148 с.
9. Dunis C., Williams M. Modelling and Trading the EUR/USD Exchange Rate: Do Neural Network Models Perform Better? //Derivatives use, trading and regulation. – 2002. – Т. 8. – №. 3. – P. 211-239.
10. Hassani H. et al. Forecasting the price of gold //Applied Economics. – 2015. – Т. 47. – №. 39. – P. 4141-4152.
11. Hossain, Mohammad. (1998). AIC and BIC – The two competitive information criteria for model selection in economics and statistics //Journal of Social Science: Part II. – №. 19. – P. 133-140.

12. Huang W. et al. Forecasting foreign exchange rates with artificial neural networks: A review //International Journal of Information Technology & Decision Making. – 2004. – Т. 3. – №. 01. – P. 145-165.

13. Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice //OTexts. – 2021. URL: <http://Otexts.com/fpp3> (дата обращения: 10.06.2021)

14. Islam M. R., Nguyen N. Comparison of Financial Models for Stock Price Prediction //Journal of Risk and Financial Management. – 2020. – Т. 13. – №. 8. – P. 181.

15. Wang J., Li X. A combined neural network model for commodity price forecasting with SSA //Soft Computing. – 2018. – Т. 22. – №. 16. – P. 5323-5333.

Интернет-ресурсы:

16. Код по курсовой работе [Электронный ресурс] URL: https://github.com/avbernukhova/course_work (дата обращения: 20.06.2021).

17. Критерий Колмогорова-Смирнова. [Электронный ресурс] URL: http://www.machinelearning.ru/wiki/index.php?title=Критерий_Колмогорова-Смирнова (дата обращения: 15.05.2021)

18. Финам. [Электронный ресурс] URL: <https://www.finam.ru> (дата обращения: 01.05.2021)

19. Introducing SSA for Time Series Decomposition. [Электронный ресурс] URL: <https://www.kaggle.com/jdarcy/introducing-ssa-for-time-series-decomposition> (дата обращения: 01.05.2021)

20. Investgazeta. Цена открытия и цена закрытия. [Электронный ресурс] URL: <https://investgazeta.ua/slovar/item/2087-tsena-otkrytiya-i-tsena-zakrytiya> (дата обращения: 01.06.2021)

21. Matplotlib. [Электронный ресурс] URL: <https://matplotlib.org> (дата обращения: 01.05.2021)

22. Plotly. [Электронный ресурс] URL: <https://plotly.com> (дата обращения: 01.05.2021)

23. Sktime. [Электронный ресурс] URL: <https://www.sktime.org/en/latest/index.html> (дата обращения: 15.06.2021).