

Математическое моделирование
Mathematical modeling

УДК 330.4

<https://doi.org/10.32362/2500-316X-2024-12-3-93-103>

EDN YSWUJG



НАУЧНАЯ СТАТЬЯ

Анализ подходов к определению тренда в структуре временного ряда

У.С. Мохнаткина,
Д.В. Парфенов,
Д.А. Петрусевич[@]

МИРЭА – Российский технологический университет, Москва, 119454 Россия

[@] Автор для переписки, e-mail: petrusevich@mirea.ru, petrdenis@mail.ru

Резюме

Цели. Основная цель – сравнить качество прогнозирования моделей временных рядов, по-разному описывающих тренд, и сформировать заключение о применимости каждого подхода при описании тренда в зависимости от свойств временного ряда.

Методы. Тренд может рассматриваться как склонность рассматриваемой величины к возрастанию или убыванию в долгосрочной перспективе. Также встречается подход, при котором тренд является функцией некоторого вида, отражающей закономерности в поведении рассматриваемого временного ряда (речь идет о закономерностях, характеризующих поведение ряда для всего рассматриваемого периода, а не краткосрочные особенности). В работе рассматривается разложение STL, построение моделей ARIMA, использование моделей ACD (усредненного условного смещения) и другие подходы. Хотя разложение на тренд, сезонность, остаток и является общеупотребительной практикой, многие комбинации, представленные в вычислительном эксперименте, построены впервые (например, использование ряда Фурье для моделирования тренда, совмещение модели сезонности и модели тренда на основе алгоритма ACD). Во второй части работы представлен вычислительный эксперимент, в котором модели, использующие различные подходы к понятию тренда, его выделению и обработке, сравниваются по значению функции максимального правдоподобия и по прогнозу на тестовый период для динамических рядов макроэкономической статистики РФ; цены акций Сбербанка РФ на Московской бирже временного периода 2000–2021 гг.

Результаты. Во всех экспериментах один из наиболее точных прогнозов сделан при помощи метода LOESS. Для сезонных рядов достаточно точные результаты показывает моделирование тренда на основе многочлена и сезонности на основе функций ARIMA, совмещение модели тренда на основе алгоритма ACD и сезонности на основе ETS и моделирование на основе ряда Фурье.

Выводы. Метод LOESS для групп сезонных и несезонных рядов дает наилучший результат по всем показателям, поэтому можно рекомендовать именно этот метод для получения наиболее точных результатов для рядов различной природы. Моделирование тренда с помощью разложения в ряд Фурье приводит к достаточно точным результатам на временных рядах различной природы. Для сезонных рядов один из лучших результатов дает комбинация моделирования тренда на основе многочлена и сезонности в виде модели ARIMA.

Ключевые слова: динамические ряды, макроэкономическая статистика, ARIMA, ACD, временные ряды, тренд, функция максимального правдоподобия, моделирование тренда

• Поступила: 21.06.2023 • Доработана: 08.02.2024 • Принята к опубликованию: 09.04.2024

Для цитирования: Мохнаткина У.С., Парфенов Д.В., Петрусеви́ч Д.А. Анализ подходов к определению тренда в структуре временного ряда. *Russ. Technol. J.* 2024;12(3):93–103. <https://doi.org/10.32362/2500-316X-2024-12-3-93-103>

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Analysis of approaches to identification of trend in the structure of the time series

Ulyana S. Mokhnatkina,
Denis V. Parfenov,
Denis A. Petrusevich @

MIREA – Russian Technological University, Moscow, 119454 Russia

@ Corresponding author, e-mail: petrusevich@mirea.ru, petrdenis@mail.ru

Abstract

Objectives. The study set out to compare the forecasting quality of time series models that describe the trend in different ways and to form a conclusion about the applicability of each approach in describing the trend depending on the properties of the time series.

Methods. A trend can be thought of as the tendency of a given quantity to increase or decrease over the long term. There is also an approach in which a trend is viewed as some function, reflecting patterns in the behavior of the time series. In this case, we discuss the patterns that characterize the behavior of the series for the entire period under consideration, rather than short-term features. The experimental part involves STL decomposition, construction of ARIMA models (one of the stages of preparation for which includes differentiation, i.e., removal of the trend and transition to a weakly stationary series), construction of ACD models (average conditional displacement) and other approaches. Time-series models based on various trend models are compared with respect to the value of the maximum likelihood function. Many of the combinations have not been constructed before (Fourier series as a trend model, combination of ACD model for trend with seasonal models). Example forecasts of macroeconomic statistics of the Russian Federation and stock prices of Sberbank on the Moscow Exchange in the time range of 2000–2021 are presented.

Results. In the experiments, The LOESS method obtained the best results. A combination of polynomial model for trend description and ARIMA for seasonally description and combination of ACD algorithm for trend and ETS for seasonal model obtained good forecasts in case of seasonal time series, while Fourier time series as a trend model also achieved close quality of prediction.

Conclusions. Since the LOESS method for groups of seasonal and non-seasonal series gives the best results for all indicators, this method can be recommended for obtaining the most accurate results for series of different nature. Trend modeling using Fourier series decomposition leads to quite accurate results for time series of different natures. For seasonal series, one of the best results is given by the combination of modeling a trend on the basis of a polynomial and seasonality in the form of the ARIMA model.

Keywords: dynamic series, macroeconomic statistics, ARIMA, ACD, time series, trend, maximum likelihood function, trend modeling

• Submitted: 21.06.2023 • Revised: 08.02.2024 • Accepted: 09.04.2024

For citation: Mokhnatkina U.S., Parfenov D.V., Petrusevich D.A. Analysis of approaches to identification of trend in the structure of the time series. *Russ. Technol. J.* 2024; 12(3):93–103. <https://doi.org/10.32362/2500-316X-2024-12-3-93-103>

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

В рамках представленной работы анализируются различные подходы к определению тренда временного ряда, существующие в современной науке. Следует отметить, что понятие тренда может быть задано с помощью нескольких определений, единого подхода к определению тренда не существует. Способы выделения тренда могут зависеть как от характеристик ряда, так и от выбранного исследователем подхода. Под трендом обычно понимают возрастание или убывание временного ряда на прогнозируемом участке. В медицине тренд, в основном, рассматривается как общее направление изменения среднего уровня характеристик в наборе данных [1]. Это определение можно обобщить до наличия постоянного однонаправленного изменения исследуемой величины [2–4]. Важно отметить наличие шумовой составляющей временного ряда, способной влиять на значения ряда как в меньшую, так и в большую сторону. По этой причине исследователей интересует компонента долгосрочного тренда, т.е. характеристики изменения исследуемой величины в течение длительного периода времени [5–8]. Также встречается способ выделения тренда в функциональном виде, что является анализом процесса на более глубоком уровне, где используется та же терминология.

В рамках работы рассмотрено несколько методов выделения тренда, приведены прогнозы на тестовый период моделей временных рядов (для случаев, где модель позволяет сделать прогноз), проведено сравнение различных подходов по качеству прогноза на тестовый период. В качестве обрабатываемых данных используются временные ряды макроэкономической статистики РФ¹, а также данные курса акций Сбербанка РФ на Московской бирже². Все рассматриваемые модели настраивались на периоде

обучения 2000–2020 гг. (из него удалены кризисные годы 2008 и 2014, данные склеены). Модели, исследуемые в работе, не зависят от конкретного периода времени, и выводы можно распространить и на другие временные процессы.

Исследователь обычно имеет под рукой множество моделей различной природы (на основе нейронных сетей, стандартные ARIMA³, ETS⁴, GARCH⁵ и т.п.), каждая из которых делает свой прогноз на целевой период. Модели, построенные в вычислительном эксперименте, могут использоваться (как показывает точность их прогнозов на тестовый период) в общем наборе предсказателей. Представленный подход актуален в связи с возможностью построить не применяемые ранее комбинации моделей для описания тренда и сезонной составляющей (если она есть) временного ряда. Это позволит добавить в существующий комплекс применяемых моделей и методов новые и объяснить, в каких случаях они могут лучше описывать динамику временного процесса по сравнению с другими моделями. Борьба за улучшение качества прогноза временного ряда остается важной вне зависимости от широты инструментария исследователя.

Цель работы состоит в построении различных моделей тренда, сопоставлении точности их прогнозов на тестовый период со стандартными моделями и построении выводов о возможности дальнейшего использования различных моделей для тренда при прогнозировании временных рядов.

В рамках концепции STL (англ., seasonal trend decomposition based on locally estimated scatterplot smoothing) [9] тренд T_t – это детерминированная часть временного ряда y_t , который может содержать также сезонную составляющую S_t и шумовую составляющую R_t . Ряд можно представить в аддитивной (1) или мультипликативной (2) форме:

¹ Единый архив экономических и социологических данных. Динамические ряды макроэкономической статистики РФ. Индекс денежных доходов населения; реального объема сельскохозяйственного производства. <http://sophist.hse.ru/hse/nindex.shtml>. Дата обращения 03.11.2023. [Unified archive of economic and sociological data. Dynamic series of macroeconomic statistics of the Russian Federation. Index of money incomes of the population; real volume of agricultural production. <http://sophist.hse.ru/hse/nindex.shtml> (in Russ.). Accessed November 03, 2023.].

² Курс акций ПАО Сбербанк (SBER). <https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER>. Дата обращения 03.11.2023. [Sberbank (SBER) stock price. <https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER> (in Russ.). Accessed November 03, 2023.].

³ Autoregressive integrated moving average model – модель авторегрессии и скользящего среднего или модель Бокса – Дженкинса.

⁴ Exponential time smoothing – экспоненциальное сглаживание.

⁵ Generalized autoregressive conditional heteroscedastic – авторегрессионная условная гетероскедастичность.

$$y_t = S_t + T_t + R_t, \quad (1)$$

$$y_t = S_t \times T_t \times R_t, \quad (2)$$

где t – время.

Подходы (1) и (2) эквивалентны, если возможно произвести замену переменных и преобразование единиц измерения на основе логарифмирования [9]:

$$\ln y_t = \ln(S_t \times T_t \times R_t) = \ln S_t + \ln T_t + \ln R_t.$$

Для выяснения наличия тренда можно использовать подход на основе статистических гипотез, в частности, критерий Манна – Кендала [10, 11].

Методы оценки тренда временного ряда делятся на параметрические и непараметрические. В параметрических методах под трендом понимают функцию одной или нескольких переменных. Примерами параметрических методов являются способы оценки тренда в виде функции, где параметры, характеризующие тренд, можно вычислять с помощью метода наименьших квадратов, полиномиальной подгонки, логарифмической, степенной, экспоненциальной, гармонической, логистической, кусочно-линейной функций, модели авторегрессии (модель AR^6) и т.д. К непараметрическим методам можно отнести модель скользящего среднего (модель MA^7), метод медианного сглаживания, метод экспоненциального сглаживания (модель ETS), применение частотных фильтров и т.д. В экспериментальной части работы рассматривается подбор логарифмической, линейной и экспоненциальной функций для описания тренда, а также применение байесовского подхода [12]. Они описывают монотонные участки роста и падения исследуемой величины, поэтому данные методы возможно применять только для монотонного временного ряда. В случае, когда исходный ряд не является монотонным, требуется произвести предварительную обработку, разбив ряд на монотонные части, и описывать каждую часть отдельно [13–17].

В экспериментальной части работы также исследуется метод описания тренда с помощью полиномов второго порядка и выше. Функция многочлена в общем случае не является монотонной, что потенциально приводит к вычислительным проблемам, а также встает вопрос выбора конкретного вида функции (степени многочлена). Данная проблема свойственна также сплайн-интерполяции. Кроме прочего, при полиномиальной интерполяции функция может сильно отклоняться от фиксированных значений в узлах. Например, известна проблема функции $1/(1+x^2)$ [18].

Рассмотрим алгоритм усредненного условного смещения ACD (average conditional displacement) [15] для оценки тренда и решения задач на его основе [12, 13]. По сравнению со сплайн-интерполяцией [19, 20] для описания поведения некоторой функции на заданном отрезке алгоритм ACD и родственные ему обладают полезным свойством сохранения монотонности. Схожие идеи также находят применение в работах на основе других подходов [14].

Кроме указанных выше способов описания тренда, широко применяются математические модели временных рядов, построенные на основе нейронных сетей (LSTM⁸, GRU⁹ и т.д.) [21–25]. Часто такие модели лучше описывают долгосрочные закономерности в данных по сравнению с моделями ARIMA, основанными на статистике [24, 26]. Известны работы, где модели разных типов интегрируются в единый механизм [22, 23]. Также в настоящее время исследователи делают попытки анализировать с помощью нейронных сетей характеристики и особенности временного ряда [26].

В экспериментальной части работы представлены построение моделей тренда на основе различных методов, объединение моделей с информацией о сезонности процесса, проверка качества прогнозов и соответствующие выводы о способности полученной модели настраиваться на значения временного ряда и о качестве прогноза на тестовый период.

РАССМАТРИВАЕМЫЕ МЕТОДЫ ВЫДЕЛЕНИЯ ТРЕНДА

Одна из самых распространенных моделей для описания временного ряда – ARIMA(p, d, q) [9], состоящая из авторегрессионной части AR (для модели порядка p значения ряда X ставятся в зависимость от p своих предыдущих значений):

$$X_t = c + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p},$$

где $\varphi_i, i = \overline{1, p}$ – коэффициенты функции; и из части скользящего среднего MA порядка q [9]:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

где $\theta_i, i = \overline{1, q}$ – коэффициенты функции. Порядок d обозначает число дифференцирований ряда.

Важно отметить, что при построении модели от тренда избавляются, переходя к стационарной временной разности (многократное дифференцирование ряда, пока не выполнится тест

⁶ Autoregressive model.

⁷ Moving average.

⁸ Long-short term memory – долгосрочная короткая память.

⁹ Gated recurrent unit – управляемый рекуррентный блок.

на стационарность) [9, 27]. Таким образом, описание модели и процесс прогнозирования производятся с преобразованным стационарным временным рядом. Модель ARIMA остается одной из самых распространенных в области прогнозирования временных рядов. В вычислительной части исследования ее результаты сравниваются с прогнозами других моделей.

В алгоритме усредненного условного смещения [15] для оценки тренда и решения задач на его основе [16, 17] участки временного ряда приближаются монотонными функциями вида:

$$f(x) = y_l + k(x - x_l), \quad (3)$$

где y_l – значение функции $f(x)$ в крайнем левом положении сегмента при $x = x_l$, k – коэффициент наклона прямой $f(x)$ (рисунок). Так как оценка тренда строится с помощью последовательного вычисления монотонных сегментов, то не возникает упомянутых проблем, как в случае полиномиальной интерполяции или сплайн-интерполяции.

В рамках вычислительного эксперимента качество прогнозов моделей ARIMA и ACD сравнивается с результатами моделей на основе нейронных сетей (LSTM, GRU) и моделей с применением беггинга [28] к данным временных рядов. Беггинг подразумевает формирование множества псевдовыборок на основе данных ряда. Итоговый прогноз получается с помощью усреднения или взвешенного усреднения прогнозов на тестовый период, построенных для каждой псевдовыборки [28].

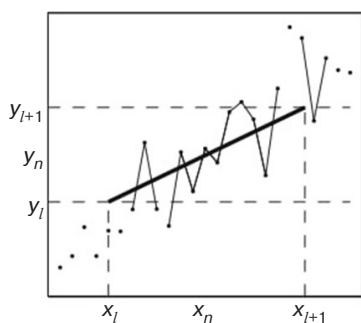


Рисунок. Схема построения монотонного сегмента по отрезку временного ряда согласно алгоритму ACD [15]

Для оценки прогнозов моделей рассматриваются меры близости вектора прогноза и вектора реальных значений временного ряда [9]:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum (\tau(t) - ts(t))^2}{N}}, \\ \text{MAE} &= \frac{\sum |\tau(t) - ts(t)|}{N}. \end{aligned} \quad (4)$$

Здесь RMSE (root mean square error) – среднеквадратичная ошибка; MAE (mean absolute error) – средняя абсолютная ошибка; $\tau(t)$ – реальные значения временного ряда; $ts(t)$ – прогноз математической модели; N – длина отрезка прогнозирования (часто совпадает с длиной интервала сезонности. В основном, речь идет о временных рядах с годичной сезонностью, поэтому $N = 12$ месяцев).

ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

В эксперименте строятся и сравниваются модели временного ряда. Для моделирования используются ряды: денежных доходов населения, реального объема сельскохозяйственного производства по данным макроэкономической статистики РФ (ежемесячные показатели, безразмерные) и акции Сбербанка на московской бирже (ежемесячные показатели, рубли). Последний год в данных используется как тестовый период, на который все модели делают прогноз. Оставшаяся часть данных используется для обучения и настройки моделей. Модели временных рядов, которые участвуют в эксперименте, используют только данные временного ряда для настройки, не учитывая внешние факторы, поэтому перед настройкой моделей мы исключили данные вокруг кризисов 2008 и 2014 гг. Данные предыдущего и следующего периодов относительно кризисного года склеиваются. Графики ряда денежных доходов, его автокорреляционной (autocorrelation function, ACF) и частичной автокорреляционной функций (partial autocorrelation function, PACF) [9] представлены и подробно описаны в [29, рис. 3 и 4]. Графики для ряда реального объема сельскохозяйственного производства представлены и подробно описаны в [29, рис. 5 и 6].

При сравнении моделей используется функция максимального правдоподобия, оценки MAE/RMSE, которые показывают близость прогноза к реальным данным тестового периода. Результаты обработки индекса денежных доходов населения представлены в табл. 1 (жирным шрифтом отмечены лучшие модели по различным критериям).

Первые пять моделей тренда представляют собой многочлены, коэффициенты которых оцениваются с помощью метода наименьших квадратов. На прогноз накладывается модель сезонности. Следующая пятерка моделей отличается тем, что прогноз по построенной полиномиальной модели тренда делается с помощью ARIMA.

При прогнозировании с помощью гиперболической/показательной/логарифмической функции тренд моделируется с помощью соответствующей функции. На основе полученной регрессионной модели тренд прогнозируется на тестовый период с последующим наложением информации о сезонности.

Таблица 1. Модели денежных доходов населения по данным макроэкономической статистики РФ и их прогнозы на тестовый период

Оценка тренда	LLF	MAE	RMSE
Многочлен 1 степени $y = a_0 + a_1x$	-364.650	0.192	0.193
Многочлен 2 степени $y = a_0 + a_1x + a_2x^2$	-363.084	0.042	0.046
Многочлен 3 степени $y = a_0 + \sum_{i=1}^3 a_i x^i$	-362.139	0.055	0.062
Многочлен 4 степени $y = a_0 + \sum_{i=1}^4 a_i x^i$	-362.746	0.016	0.020
Многочлен 5 степени $y = a_0 + \sum_{i=1}^5 a_i x^i$	-363.487	0.081	0.086
Многочлен 1 степени + ARIMA	-364.151	0.145	0.147
Многочлен 2 степени + ARIMA	-362.863	0.024	0.027
Многочлен 3 степени + ARIMA	-362.453	0.026	0.033
Многочлен 4 степени + ARIMA	-362.668	0.014	0.019
Многочлен 5 степени + ARIMA	-362.746	0.016	0.020
Гиперболическая функция $y = a_0 + a_{-1} / x$	-361.679	0.105	0.107
Логарифмическая функция $y = a_0 + a_1 \ln x$	-363.174	0.050	0.054
Экспоненциальная функция $y = \exp(a_0 + a_1 x)$	-364.967	0.221	0.222
Интерполяция сплайнами	-362.688	–	–
Разложение в ряд Фурье	-362.783	0.020	0.027
Экспоненциальное сглаживание	-363.073	0.040	0.044
Метод LOESS	-362.719	0.015	0.019
Алгоритм ACD	-362.959	0.030	0.034
Тренд ACD + ARIMA	-362.198	0.060	0.074
Тренд ACD + ETS	-361.777	0.010	0.012

Коэффициенты модели также вычисляются с помощью метода наименьших квадратов.

Разложение в ряд Фурье используется для описания модели тренда на обучающий период. Прогнозирование осуществляется с помощью ARIMA аналогично представленным выше методам с наложением модели сезонности.

Интерполяцию функции, заданной таблично (что, в целом, соответствует измерениям, по которым строится временной ряд), можно провести с помощью сплайнов. Интересно, что такой метод интерполяции обладает высокой точностью, но на его основе не удастся сделать прогноз. При этом можно сравнить то, насколько сплайны и другие модели подстроены под данные ряда.

При использовании модели STL (метод LOESS¹⁰) [30, 31] ряд делится на составляющие: тренд, сезонность, шум.

При применении алгоритма ACD подход аналогичен. Модель тренда строится на основе ACD, производится прогноз на тестовый период.

Также исследуются гибридные модели, в которых на модель тренда накладывается сезонность и делается прогноз с помощью ETS или ARIMA.

Вычислительный эксперимент следует методике, используемой в работе [29]. В первой части используется временной ряд ежемесячных показателей

¹⁰ Locally estimated scatterplot smoothing – метод локального сглаживания диаграммы рассеивания.

денежных доходов населения. Ряд имеет годовую сезонность. Тестовым периодом является 2018 год. Результаты приведены в табл. 1 (точность – 0.01). Перед проведением эксперимента данные ряда были преобразованы в диапазон [0, 1]. Лучшие модели по точности прогноза на тестовый период или по значению логарифма функции максимального правдоподобия (LLF, log-likelihood function) выделены жирным шрифтом.

Самые точные результаты (столбцы MAE и RMSE) получаются при рассмотрении гибридной модели с оценкой тренда алгоритмом ACD с информацией о сезонности временного ряда и ее прогнозом на тестовый период с помощью ETS. Хорошие показатели по точности прогноза на тестовый период у метода оценки тренда с помощью многочлена четвертой степени с добавлением информации о сезонности и случайной составляющей на основе модели ARIMA(0, 4, 1). Значения функции правдоподобия меньше по сравнению с предыдущим методом. Такие же значения ошибок и качества модели для оценки тренда с помощью метода LOESS. Данные модели имеют почти те же показатели точности подстройки под данные ряда (LLF), что и при моделировании ряда на основе сплайнов.

Следующий по показателям точности результатов и качеству подстройки под исходный ряд – способ оценки тренда с помощью многочлена четвертой степени, смоделированного и спрогнозированного методом линейной регрессии. Такие же показатели имеет метод оценки тренда с помощью многочлена пятой степени с прогнозом на тестовый период методом ARIMA(3, 5, 1).

Неплохие оценки по точности прогноза и по функции правдоподобия при моделировании тренда с помощью разложения в ряд Фурье, многочленов второй и третьей степени с прогнозированием с помощью метода ARIMA. Чуть ниже точность прогноза при моделировании тренда с помощью алгоритма ACD.

Классические методы ARIMA и ETS имеют менее точные показатели прогноза на тестовый период, но далеко не худшие.

Наиболее низкие значения для логарифма функции максимального правдоподобия имеют методы оценки тренда с помощью экспоненциальной функции, а также с помощью многочлена первой степени. Однако данные модели имеют худшие результаты точности прогноза на тестовый период. Худший по всем показателям – метод моделирования тренда с помощью гиперболической функции.

Аналогичный вычислительный эксперимент проведен для временного ряда ежемесячных показателей индекса реального объема сельскохозяйственного производства (подробный анализ характеристик ряда приведен в работе [29]). Данные

2000–2020 гг. используются для обучения. Ряд имеет годовую сезонность. Тестовым периодом является 2021 год. Результаты приведены в табл. 2. Лучшие модели по точности прогноза на тестовый период или по значению логарифма функции максимального правдоподобия выделены жирным шрифтом.

Самые низкие значения для логарифма функции максимального правдоподобия имеет гибридная модель с оценкой тренда алгоритмом ACD с информацией о сезонности временного ряда и ее прогнозом на тестовый период с помощью ETS. Однако данный метод имеет не самые точные результаты прогнозов на тестовый период (столбцы MAE и RMSE).

Лучшие результаты по показателю MAE для прогноза на тестовый период у метода оценки тренда с помощью логарифмической и экспоненциальной функции (лучшие модели выделены жирным шрифтом). Близкие к ним значения у моделей с разложением в ряд Фурье, экспоненциальным сглаживанием, гиперболической функцией, а также с методом LOESS.

Лучшие результаты по показателю RMSE для прогноза на тестовый период у метода оценки тренда с помощью разложения в ряд Фурье. Также неплохие значения ошибок у моделирования тренда с помощью экспоненциальной функции, экспоненциального сглаживания, метода LOESS и с помощью многочлена первой степени с прогнозированием как на основе ARIMA, так и на основе линейной регрессии.

Наименьшее значение функции правдоподобия после гибридной модели с трендом ACD и прогнозом ETS имеет способ моделирования тренда с помощью многочлена пятой степени с прогнозированием линейной регрессией. Немного больше данное значение для гибридной модели с трендом ACD и информацией о сезонности временного ряда, спрогнозированной ARIMA. Также можно отметить модели с оценкой тренда с помощью многочленов третьей и четвертой степени. Однако данная группа моделей имеет одни из худших показателей для точности прогноза.

Важно отметить, что при оценке тренда с помощью многочленов точность снижается с ростом степени многочлена. Данное свойство сохраняется и при объединении оценки тренда с помощью многочленов с моделью ARIMA.

В представленных выше примерах обрабатываются сезонные временные ряды. Рассмотрим работу различных методов оценки тренда на несезонном временном ряду акций Сбербанка РФ¹¹ (руб.).

¹¹ Курс акций ПАО Сбербанк (SBER). <https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER>. Дата обращения 03.11.2023. [Sberbank (SBER) stock price. <https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER> (in Russ.). Accessed November 03, 2023.].

Таблица 2. Модели индекса реального объема сельскохозяйственного производства по данным макроэкономической статистики РФ и их прогнозы на тестовый период

Оценка тренда	LLF	MAE	RMSE
Многочлен 1 степени $y = a_0 + a_1x$	-395.597	0.084	0.097
Многочлен 2 степени $y = a_0 + a_1x + a_2x^2$	-395.871	0.096	0.105
Многочлен 3 степени $y = a_0 + \sum_{i=1}^3 a_i x^i$	-396.132	0.112	0.117
Многочлен 4 степени $y = a_0 + \sum_{i=1}^4 a_i x^i$	-396.173	0.114	0.119
Многочлен 5 степени $y = a_0 + \sum_{i=1}^5 a_i x^i$	-397.175	0.171	0.185
Многочлен 1 степени + ARIMA	-395.567	0.083	0.096
Многочлен 2 степени + ARIMA	-395.734	0.090	0.100
Многочлен 3 степени + ARIMA	-395.830	0.093	0.103
Многочлен 4 степени + ARIMA	-395.836	0.094	0.103
Многочлен 5 степени + ARIMA	-396.031	0.102	0.107
Гиперболическая функция $y = a_0 + a_{-1} / x$	-394.600	0.079	0.124
Логарифмическая функция $y = a_0 + a_1 \ln x$	-395.058	0.069	0.101
Экспоненциальная функция $y = \exp(a_0 + a_1 x)$	-395.310	0.075	0.095
Интерполяция сплайнами	-396.028	–	–
Преобразования Фурье	-395.648	0.076	0.089
Экспоненциальное сглаживание	-395.343	0.078	0.096
Метод LOESS	-395.379	0.079	0.097
Алгоритм ACD	-396.643	0.145	0.153
Тренд ACD + ARIMA	-397.140	0.170	0.195
Тренд ACD + ETS	-400.076	0.121	0.121

Данные 2000–2021 гг. используются для обучения. Тестовым периодом является 2022 год. Графики функций ACF и PACF представлены и подробно описаны в [29, рис. 7]. Результаты моделирования и прогнозирования представлены в табл. 3. Лучшие модели по точности прогноза на тестовый период или по значению логарифма функции максимально-го правдоподобия выделены жирным шрифтом.

Для биржевых данных сезонность отсутствует, поэтому результат прогноза зависит полностью от модели тренда и случайной составляющей.

Наилучшим образом поведение ряда описывают модели функции сплайнов и метода LOESS. При хороших показателях функции правдоподобия модель тренда с помощью метода LOESS имеет самые

точные прогнозы. Ошибка составляет 0.004 для MAE и 0.006 для RMSE. Также низкие значения LLF соответствуют моделированию тренда с помощью многочлена пятой степени с прогнозом линейной регрессии и с помощью разложения в ряд Фурье. Данные модели также имеют одни из лучших показателей точности после модели метода LOESS.

Остальные модели, в отличие от моделирования сезонных рядов, хуже подстраиваются под поведение ряда. Для них характерно ухудшение точности по мере возрастания значений LLF.

При моделировании тренда с помощью многочлена при возрастании степени многочлена, начиная с четвертой степени, точность растет, что отчасти может быть связано с переобучением. Это относится

Таблица 3. Модели временного ряда акций Сбербанка РФ и их прогнозы на тестовый период

Оценка тренда	LLF	MAE	RMSE
Многочлен 1 степени $y = a_0 + a_1x$	-4979.271	0.105	0.128
Многочлен 2 степени $y = a_0 + a_1x + a_2x^2$	-4950.327	0.199	0.215
Многочлен 3 степени $y = a_0 + \sum_{i=1}^3 a_i x^i$	-4881.779	0.440	0.465
Многочлен 4 степени $y = a_0 + \sum_{i=1}^4 a_i x^i$	-4986.211	0.108	0.114
Многочлен 5 степени $y = a_0 + \sum_{i=1}^5 a_i x^i$	-5019.458	0.079	0.119
Многочлен 1 степени + ARIMA	-4993.657	0.070	0.093
Многочлен 2 степени + ARIMA	-4995.075	0.067	0.089
Многочлен 3 степени + ARIMA	-4997.347	0.063	0.083
Многочлен 4 степени + ARIMA	-5006.537	0.067	0.085
Многочлен 5 степени + ARIMA	-5006.588	0.062	0.079
Гиперболическая функция $y = a_0 + a_{-1} / x$	-4880.421	0.443	0.451
Логарифмическая функция $y = a_0 + a_1 \ln x$	-4923.877	0.288	0.300
Экспоненциальная функция $y = \exp(a_0 + a_1 x)$	-4992.214	0.072	0.094
Интерполяция сплайнами	-5013.558	—	—
Преобразования Фурье	-5014.470	0.038	0.054
Экспоненциальное сглаживание	-4976.059	0.117	0.145
Метод LOESS	-5013.578	0.004	0.006
Алгоритм ACD	-4996.171	0.077	0.086
Тренд ACD + ARIMA	-4987.216	0.145	0.162
Тренд ACD + ETS	-4979.289	0.170	0.191

и к комбинации моделирования ряда с помощью многочлена и ARIMA, но показатели точности улучшаются при росте степени многочлена, начиная с первой степени.

ЗАКЛЮЧЕНИЕ

При моделировании тренда для сезонных рядов наложение на него информации о сезонной и случайной составляющих влияет на качество прогнозов. Лучшие результаты показывают способы моделирования тренда с помощью разложения в ряд Фурье и метод LOESS. Также достаточно точные результаты у комбинации моделирования тренда с помощью многочлена с методом ARIMA для сезонности.

При использовании многочленов для оценки тренда с прогнозом линейной регрессии показатели хуже, чем при использовании комбинации многочлена с моделью ARIMA. При этом динамика поведения показателей точности у них одинаковая.

Интересно заметить, что алгоритм ACD лучше всего проявил себя на данных денежных доходов населения. Этот временной ряд имеет неоднородную дисперсию. Прогноз данных с использованием алгоритма ACD может быть весьма полезен для гетероскедастических рядов.

Моделирование тренда с помощью экспоненциальной и логарифмической функций не продемонстрировало выдающихся результатов. Данные методы также сложнее с вычислительной точки зрения

по сравнению с моделями многочленов. Модель логарифмической функции имеет ограничения на значения данных, т.к. вещественного логарифма отрицательного аргумента не существует. Модель гиперболической функции является одной из худших как по показателям функции правдоподобия, так и по оценкам точности.

В отличие от оценки тренда с помощью экспоненциальной функции, выделение тренда с использованием экспоненциального сглаживания привело к одному из лучших результатов. Недостатком данного метода является неопределенный коэффициент сглаживания.

При работе с несезонными временными рядами качество прогноза зависит лишь от трендовой и шумовой составляющих. Лучше всего под данные подстраиваются модели с разложением в ряд Фурье,

методом LOESS и функцией сплайнов. Однако на основе сплайнов трудно сделать прогноз вследствие их ориентации на интерполяцию данных.

Следует отметить, что метод LOESS для группы несезонных рядов также дает наилучший или близкий к лучшему результат по всем показателям, поэтому можно рекомендовать именно этот метод получения наиболее точных результатов для рядов различной природы. Также можно отметить моделирование тренда с помощью разложения в ряд Фурье, т.к. данный подход также приводит к достаточно точным результатам на временных рядах различной природы.

Вклад авторов. Все авторы в равной степени внесли свой вклад в исследовательскую работу.

Authors' contribution. All authors equally contributed to the research work.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Allen R. Time series methods in the monitoring of intracranial pressure. Part 1: Problems, suggestion for a monitoring scheme and review of appropriate techniques. *J. Biomed. Eng.* 1983;5(1):5–18. [https://doi.org/10.1016/0141-5425\(83\)90073-0](https://doi.org/10.1016/0141-5425(83)90073-0)
2. Blom J.A., Ruyter J.F., Saranummi F., Beneken J.W. Detection of trends in monitored variables. In: Carson E.R., Cramp D.G. (Eds.). *Computer and Controls in Clinical Medicine*. New York: Plenum; 1985. P. 153–174. https://doi.org/10.1007/978-1-4613-2437-9_6
3. Challis R.E., Kitney R.I. Biomedical signal processing (in four parts). Part I: Time domain methods. *Med. Biol. Eng. Comput.* 1990;28(6):509–524. <https://doi.org/10.1007/bf02442601>
4. Haimowitz I.J., Kohane I.S. Automated trend detection with alternative temporal hypotheses. In: *Proceedings of the 13th International Joint Conference of Artificial Intelligence IJCAI-93*. 1993. P. 146–151.
5. Helsel D.R., Hirsch R.M., Ryberg K.R., Archfield S.A. *Statistical Methods in Water Resources*. USGS Science Publishing Network, Reston Publishing Service Center; 2018. 458 p. ISBN 978-1-4113-4348-1. <https://doi.org/10.3133/tm4a3>
6. Ding H., Li Z., Ren Q., Chen H., Song M., Wang Y. Single-variable method for predicting trends in chlorophyll a concentration based on the similarity of time series. *Ecological Indicators*. 2022;14096):109027. <https://doi.org/10.1016/j.ecolind.2022.109027>
7. Yao J., Wang P., Wang G., Shrestha S., Xue B., Sun W. Establishing a time series trend structure model to mine potential hydrological information from hydrometeorological time series data. *Sci. Total Environ.* 2020;698:134227. <https://doi.org/10.1016/j.scitotenv.2019.134227>
8. De Leo F., De Leo A., Besio G., Briganti R. Detection and quantification of trends in time series of significant wave heights: An application in the Mediterranean Sea. *Ocean Eng.* 2020;202:107155. <https://doi.org/10.1016/j.oceaneng.2020.107155>
9. Hyndman R.J., Athanasopoulos G. *Forecasting: Principles and Practice*. 3rd ed. OTexts; 2021. 442 p. ISBN-13 978-0-9875-0713-6
10. Mann H.B. Nonparametric tests against trend. *Econometrica*. 1945;13(3):2453–259. <https://doi.org/10.2307/1907187>
11. Kendall M.G. *Rank Correlation Methods*. 2nd ed. Hafner Publishing Co.; 1955. 196 p.
12. Kohns D., Bhattacharjee A. Nowcasting growth using Google Trends data: A Bayesian Structural Time Series model. *Int. J. Forecast.* 2022;39(3):1384–1412. <https://doi.org/10.1016/j.ijforecast.2022.05.002>
13. Yahyaoui H., Al-Daihani R. A novel trend based SAX reduction technique for time series. *Expert Systems with Applications*. 2019;130(C):113–123. <https://doi.org/10.1016/j.eswa.2019.04.026>
14. Xie Y., Liu S., Huang S., Fang H., Ding M., Huang C., Shen T. Local trend analysis method of hydrological time series based on piecewise linear representation and hypothesis test. *J. Clean. Prod.* 2022;339(1):130695. <https://doi.org/10.1016/j.jclepro.2022.130695>
15. Vamoş C., Crăciun M. *Automatic Trend Estimation*. Dordrecht, Heidelberg, New York, London: Springer; 2013. 131 p. <https://doi.org/10.1007/978-94-007-4825-5>
16. Feng Y., Zhou C. Forecasting financial market activity using a semiparametric fractionally integrated Log-ACD. *Int. J. Forecast.* 2015;31(2):349–363. <http://doi.org/10.1016/j.ijforecast.2014.09.001>
17. Allen D., Chan F., McAleer M., Peiris S. Finite sample properties of the QMLE for the Log-ACD model: Application to Australian stocks. *J. Econometrics*. 2008;147(1):163–185. <https://doi.org/10.1016/j.jeconom.2008.09.020>

18. Epperson J. On the Runge example. *The American Mathematical Monthly*. 1987;94(4):329–341. <https://doi.org/10.2307/2323093>
19. Drozdov I., Petrushevich D. Water pollution time series analysis. *IOP Conf. Ser.: Mater. Sci. Eng.* 2021;1047(1):012095. <http://doi.org/10.1088/1757-899X/1047/1/012095>
20. Petrushevich D. Review of missing values procession methods in time series data. *J. Phys.: Conf. Ser.* 2021;1889(3):032009. <http://doi.org/10.1088/1742-6596/1889/3/032009>
21. Wang P., Zheng X., Ai G., Liu D., Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. *Chaos, Solitons & Fractals*. 2020;140:110214. <https://doi.org/10.1016/j.chaos.2020.110214>
22. Kumar B., Sunil P., Yadav N. A novel hybrid model combining β SARMA and LSTM for time series forecasting. *Appl. Soft Comput.* 2023;134:110019. <https://doi.org/10.1016/j.asoc.2023.110019>
23. Abebe M., Noh Y., Kang Y.-J., Seo C., Kim D., Seo J. Ship trajectory planning for collision avoidance using hybrid ARIMA-LSTM models. *Ocean Eng.* 2022;256:111527. <https://doi.org/10.1016/j.oceaneng.2022.111527>
24. Arunkumar K.E., Kalaga D.V., Kumar M.S., Kawaji M., Brenza T.M. Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Eng. J.* 2022;61(10):7585–7603. <https://doi.org/10.1016/j.aej.2022.01.011>
25. Ning Y., Kazemi H., Tahmasebi P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Comput. Geosci.* 2022;164:105126. <https://doi.org/10.1016/j.cageo.2022.105126>
26. Anghinoni L., Zhao L., Ji D., Pan H. Time series trend detection and forecasting using complex network topology analysis. *Neural Netw.* 2019;117:295–306. <https://doi.org/10.1016/j.neunet.2019.05.018>
27. Box G., Jenkins G., Reinsel G.C. *Time Series Analysis: Forecasting and Control*. John Wiley and Sons; 2008. 784 p. ISBN-13 978-0470272848
28. Petropoulos F., Hyndman R.J., Bergmeir C. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *Eur. J. Oper. Res.* 2018;268(2):545–554. <https://doi.org/10.1016/j.ejor.2018.01.045>
29. Грамович Я.В., Мусатов Д.Ю., Петрусеви Д.А. Применение беггинга в прогнозировании временных рядов. *Russ. Technol. J.* 2024;12(1):101–110. <https://doi.org/10.32362/2500-316X-2024-12-1-101-110>
[Gramovich I.V., Musatov D.Yu., Petrushevich D.A. Implementation of bagging in time series forecasting. *Russ. Technol. J.* 2024;12(1):101–110. <https://doi.org/10.32362/2500-316X-2024-12-1-101-110>]
30. Zhao K., Wulder M.A., Hu T., Bright R., Wu Q., Qin H., Li Y., Toman E., Mallick B., Zhang X., Brown M. Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sens. Environ.* 2019;232:111181. <https://doi.org/10.1016/j.rse.2019.04.034>
31. Li J., Li Z.-L., Wu H., You N. Trend, seasonality, and abrupt change detection method for land surface temperature time-series analysis: Evaluation and improvement. *Remote Sens. Environ.* 2022;280:113222. <https://doi.org/10.1016/j.rse.2022.113222>

Об авторах

Мохнаткина Ульяна Станиславовна, студент, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: atlantika@live.ru. <https://orcid.org/0009-0008-8756-7267>

Парфенов Денис Васильевич, к.т.н., доцент, доцент кафедры высшей математики, Институт искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: parfenov@mirea.ru. Scopus Author ID 57217119805, SPIN-код РИНЦ 7463-3220, <https://orcid.org/0009-0004-0905-3827>

Петрусеви Денис Андреевич, к.ф.-м.н., доцент кафедры высшей математики, Институт искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: petrusevich@mirea.ru, petrdenis@mail.ru. Scopus Author ID 55900513600, ResearcherID AAA-6661-2020, SPIN-код РИНЦ 7999-6345, <https://orcid.org/0000-0001-5325-6198>

About the authors

Ulyana S. Mokhnatkina, Student, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: atlantika@live.ru. <https://orcid.org/0009-0008-8756-7267>

Denis V. Parfenov, Cand. Sci. (Eng.), Associate Professor, Higher Mathematics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: parfenov@mirea.ru. Scopus Author ID 57217119805, RSCI SPIN-code 7463-3220, <https://orcid.org/0009-0004-0905-3827>

Denis A. Petrushevich, Cand. Sci. (Phys.-Math.), Associate Professor, Higher Mathematics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: petrusevich@mirea.ru, petrdenis@mail.ru. Scopus Author ID 55900513600, ResearcherID AAA-6661-2020, RSCI SPIN-code 7999-6345, <https://orcid.org/0000-0001-5325-6198>