Mathematical modeling

Математическое моделирование

UDC 330.4
https://doi.org/10.32362/2500-316X-2024-12-3-93-103
EDN YSWUJG

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.

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., Петрусевич 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.

INTRODUCTION

Within the scope of the presented work, various approaches to the definition of the trend of a time series existing in modern science are analyzed. Since there is no single approach to the definition of a trend, the concept can be given a number of definitions depending both on the characteristics of the series and the approach chosen by the researcher. However, a trend is usually understood as an increasing or decreasing propensity of a time series in the predicted area. In medicine, trend is typically considered as the general direction of change in the average level of characteristics in a data set [1]. This definition can be generalized to the presence of a constant unidirectional change in the quantity under study [2–4]. It is important to note the presence of the noise component of the time series, which can affect the values of the series in both downward and upward directions. For this reason, researchers are interested in the component of long-term trend, i.e., the characteristics of changes in the studied value over a long period of time [5–8]. There is also a way to identify the trend in functional form, which is the analysis of the process at a deeper level, where the same terminology is used.

The article considers several methods of trend extraction, providing forecasts for the test period of time series models (for cases where the model enables making a forecast), and comparing different approaches in terms of the quality of forecasts for the test period. In terms of processed data, we used the time series of macroeconomic statistics of the Russian Federation[1], as well as data of the stack quote of Sberbank of the Russian Federation (RF Sberbank) on the Moscow Exchange[2]. All the models under consideration were tuned on the 2000–2020 study period (following removal of the crisis years 2008 and 2014, the data were joined). However, since the models studied in this paper do not depend on a specific time period, the conclusions can be extended to other time processes.

The researcher generally has at hand many models of different natures (based on neural networks, standard autoregressive integrated moving average model (ARIMA)[3], exponential time smoothing (ETS), generalized autoregressive conditional heteroscedastic (GARCH), etc.), each of which makes its own prediction for the target period. The models built in computational experiments can be used (as the accuracy of their predictions for the test period shows) in a common set of predictors. The presented approach is relevant due to the possibility to build combinations of models for describing trends and seasonal components (if any) of time series that have not been used before. This will make it possible to add new models and methods to the existing set of applied models and methods, as well as to explain when the dynamics of the time process can be better described in comparison with other models. The struggle to improve the quality of time series forecasting remains important regardless of the breadth of the researcher's toolkit.

The aim of the work is to build various trend models, to compare the accuracy of their forecasts for the test period with standard models, and to draw conclusions about the possibility of further use of various models for trend in time series forecasting.

Within the concept of STL (seasonal trend decomposition based on locally estimated scatterplot smoothing) [9], trend T_t is a deterministic part of the time series y_t, which may also contain seasonal component S_t and noise component R_t. The series can be represented in additive (1) or multiplicative (2) form:

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

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

where t is time.

If variable substitution and unit conversion based on logarithmization are possible, approaches (1) and (2) are equivalent [9]:

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

[1] 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] Sberbank (SBER) stock price. https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER (in Russ.). Accessed November 03, 2023.

[3] Autoregressive and moving average model or Box–Jenkins model.

In order to find out the presence of a trend, a statistical hypothesis approach can be used—in particular, the Mann–Kendall criterion [10, 11].

Methods for estimating the trend of a time series are divided into parametric and non-parametric methods. In terms of parametric methods, a trend is understood as a function of one or more variables. Examples of parametric methods are the methods of trend estimation in the form of a function, where the parameters characterizing the trend can be calculated using the least squares method, polynomial fitting, logarithmic, step, exponential, harmonic, logistic functions, piecewise linear function, autoregressive (AR) model, etc. Non-parametric methods include the moving average (MA) model, median smoothing method, ETS model, application of frequency filters, etc. The experimental part of the work considers the selection of logarithmic, linear and exponential functions to describe the trend, as well as the application of the Bayesian approach [12]. Since describing monotonic areas of growth and decline of the studied value, these methods can be applied only for a monotonic time series. In the case when the initial series is not monotonic, it is necessary to perform preliminary processing, dividing the series into monotonic parts, and describing each part separately [13–17].

The experimental section of the work also studies the method of describing a trend using polynomials of the second order and higher. Since the polynomial function not in generally monotonic, computational problems may arise along with the question of choosing a specific type of function (degree of the polynomial). This problem is also characteristic of spline interpolation. Among other things, in polynomial interpolation the function can deviate strongly from the fixed values at the nodes. For example, the function $1/(1+x^2)$ problem is known [18].

Let us consider the average conditional displacement ACD (average conditional displacement) algorithm [15] for trend estimation and solving problems on this basis [12, 13]. Compared to spline interpolation [19, 20] for describing the behavior of some function on a given segment, the ACD algorithm and related algorithms have the useful property of preserving monotonicity. Similar ideas also find application in works based on other approaches [14].

In addition to the above trend description methods, mathematical models of time series based on neural networks (long-short term memory (LSTM), gated recurrent unit (GRU), etc.) are widely used [21–25]. Often such models better describe long-term patterns in the data compared to ARIMA models based on statistics [24, 26]. There are known works where models of different types are integrated into a single framework [22, 23]. Researchers are currently making attempts to analyze the characteristics and features of time series using neural networks [26].

The experimental section of the work presents the construction of trend models based on different methods, combining the models with information on the seasonality of the process, checking the quality of forecasts and corresponding conclusions concerning the ability of the obtained model to adjust to the values of the time series and the quality of the forecast for the test period.

TREND EXTRACTION METHODS UNDER CONSIDERATION

One of the most commonly used models for describing a time series is ARIMA(p, d, q) [9], which consists of the autoregressive part of AR (for a model of order p the values of the series X are made dependent on p of their previous values):

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

where φ_i, $i = \overline{1, p}$ are the function coefficients; and from the moving average MA part of order q [9]:

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

where θ_i, $i = \overline{1, q}$ are the function coefficients. The order d denotes the number of differentiations of the series.

When building the model, the trend is eliminated by switching to the stationary time difference (multiple differentiation of the series until the test for stationarity is fulfilled) [9, 27]. Thus, the model description and the forecasting process are performed with the transformed stationary time series. The ARIMA model remains one of the most widely used models in the field of time series forecasting. In the computational part of the study, its results are compared with the forecasts of other models.

In the averaged conditional bias algorithm [15] for estimating a trend and solving related problems [16, 17], time series segments are approximated by monotonic functions of the following form:

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

where y_l is the value of the function $f(x)$ at the leftmost position of the segment at $x = x_l$, k is the slope coefficient of the line $f(x)$ (see Figure).

Since the trend estimation is constructed by successive calculation of monotone segments, the mentioned problems do not arise, as in the case of polynomial interpolation or spline interpolation.

As a part of the computational experiment, the quality of ARIMA and ACD model predictions was compared with the results of models based on neural networks (LSTM, GRU) and models with bagging [28] applied to time series data. Bagging involves generating a set of pseudo-samples from the series data. The final forecast is obtained by averaging or weighted averaging of forecasts for the test period constructed for each pseudo-sample [28].

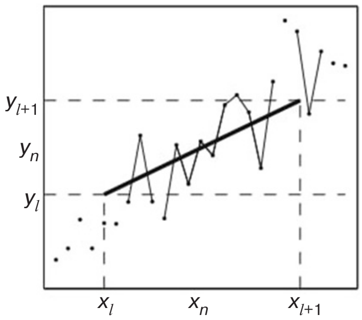

Fig. Diagram of the monotonic segment construction by time series segment according to the ACD algorithm [15]

In order to evaluate model forecasts, measures of forecast vector closeness and the vector of real values of the time series are considered [9]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_t (\tau(t) - ts(t))^2}{N}},\qquad \mathrm{MAE} = \frac{\sum_t |\,\tau(t) - ts(t)\,|}{N}. \tag{4}$$

Here RMSE is the root mean square error; MAE is the mean absolute error; $\tau(t)$ are the real values of the time series; $ts(t)$ is the forecast of the mathematical model; N is the length of the forecasting interval (often coincides with the length of the seasonality interval; since we are talking here about time series with annual seasonality, $N = 12$ months).

COMPUTATIONAL EXPERIMENT

In the experiment, time series models are built and compared. The following series are used for modeling: monetary incomes of the population, real volume of agricultural production according to macroeconomic statistics of the Russian Federation (monthly indicators, dimensionless) and stocks of RF Sberbank on the Moscow Exchange (monthly indicators, rubles). The last year in the data is used as a test period for which forecasts are made by all models. The rest of the data is used for training and tuning the models. Since the time series models that participate in the experiment use only time series data for tuning without considering external factors, we excluded data around the 2008 and 2014 crises prior to tuning the models. Data from the previous and the next period relative to the crisis year are joined together. Graphs of the money income series, as well as its autocorrelation function (ACF) and partial autocorrelation function (PACF) [9], are presented and described in detail in [29, Figs. 3 and 4]. Graphs for a series of real agricultural output are presented and described in detail in [29, Figs. 5 and 6].

The maximum likelihood function and MAE/RMSE estimates used in model comparison show the closeness of the forecast to the real data of the test period. The results of processing the index of money income of the population are presented in Table 1 (the best models according to various criteria are marked in bold).

First five trend models are polynomials whose coefficients are estimated using the least squares method. A seasonality model is superimposed on the forecast. The next five models are characterized by the fact that the forecast on the constructed polynomial trend model is made via ARIMA.

When forecasting using hyperbolic/indicative/logarithmic function, the trend is modeled using the corresponding function. Based on the obtained regression model, the trend is forecast for the test period with subsequent imposition of seasonality information. The model coefficients are also calculated using the least squares method.

Fourier series expansion is used to describe the trend model for the training period. Forecasting is performed using ARIMA similarly to the methods presented above but with imposition of the seasonality model.

Interpolation of a tabulated function (which, in general, corresponds to the measurements on which the time series is constructed) can be performed using splines. Interestingly, while such an interpolation method is highly accurate, it cannot be used to make predictions. However, it is possible to compare how splines and other models are fitted to the series data.

When using the STL model (locally estimated scatterplot smoothing (LOESS) method) [30, 31] the series is divided into components: trend, seasonality, noise.

A similar approach is taken when applying the ACD algorithm. After building the trend model on the basis of ACD, the forecast for the test period is carried out.

Hybrid models, in which seasonality is superimposed on a trend model and predicted using ETS or ARIMA, are also studied.

Computational experiment follows the methodology used in [29]. The first part uses a time series of monthly indicators of money income of the population. The series has annual seasonality. The test period is 2018. The results are summarized in Table 1 (accuracy is 0.01). Prior to the experiment, the data of the series were transformed into the range [0, 1]. The best models in terms of forecast accuracy for the test period or the value of the logarithm of the maximum log-likelihood function (LLF) are highlighted in bold type.

Table 1. Household money income models according to macroeconomic statistics of the Russian Federation and their forecasts for the test period

Trend estimation	LLF	MAE	RMSE
1st degree polynomial $y = a_0 + a_1x$	−364.650	0.192	0.193
2nd degree polynomial $y = a_0 + a_1x + a_2x^2$	−363.084	0.042	0.046
3rd degree polynomial $y = a_0 + \sum_{i=1}^{3} a_ix^i$	−362.139	0.055	0.062
4th degree polynomial $y = a_0 + \sum_{i=1}^{4} a_ix^i$	**−362.746**	**0.016**	**0.020**
5th degree polynomial $y = a_0 + \sum_{i=1}^{5} a_ix^i$	−363.487	0.081	0.086
1st degree polynomial + ARIMA	−364.151	0.145	0.147
2nd degree polynomial + ARIMA	−362.863	0.024	0.027
3rd degree polynomial + ARIMA	−362.453	0.026	0.033
4th degree polynomial + ARIMA	**−362.668**	**0.014**	**0.019**
5th degree polynomial + ARIMA	**−362.746**	**0.016**	**0.020**
Hyperbolic function $y = a_0 + a_{-1} / x$	−361.679	0.105	0.107
Logarithmic function $y = a_0 + a_1 \ln x$	−363.174	0.050	0.054
Exponential function $y = \exp(a_0 + a_1x)$	−364.967	0.221	0.222
Interpolation by splines	−362.688	–	–
Fourier series expansion	−362.783	0.020	0.027
Exponential smoothing	−363.073	0.040	0.044
LOESS method	**−362.719**	**0.015**	**0.019**
ACD algorithm	−362.959	0.030	0.034
ACD + ARIMA trend	−362.198	0.060	0.074
ACD + ETS trend	**−361.777**	**0.010**	**0.012**

The most accurate results (MAE and RMSE columns) are obtained when considering a hybrid model with trend estimation by ACD algorithm with information on seasonality of the time series and its forecast for the test period via ETS. The method of trend estimation using a polynomial of the fourth degree with the addition of seasonality information and a random component based on the ARIMA(0, 4, 1) model has good performance in terms of forecast accuracy for the test period. The values of the likelihood function are smaller as compared to the previous method. The same values of error and model quality for trend estimation using LOESS method. These models have almost the same values of accuracy of fitting to the data of the series (LLF) as the modeling of the series based on splines.

The next method of trend estimation in terms of accuracy of results and quality of adjustment to the initial series uses a polynomial of the fourth degree modeled and predicted by linear regression method. The same indicators have the method of trend estimation using a polynomial of the fifth degree with a forecast for the test period by the ARIMA(3, 5, 1) method.

Fairly good estimates of forecast accuracy and likelihood function when modeling the trend using Fourier series expansion, polynomials of the second and third degree with forecasting using the ARIMA method. Slightly lower forecast accuracy is obtained when modeling the trend using the ACD algorithm.

While classical ARIMA and ETS methods have less accurate prediction performance for the test period, they are far from useless.

The lowest values for the logarithm of the maximum likelihood function pertain to the methods of trend estimation using the exponential function, as well as those using the polynomial of the first degree. However, these models also have the worst prediction accuracy results for the test period. The worst method of trend modeling by all indicators is the one using the hyperbolic function.

A similar computational experiment was conducted for the time series of monthly indicators of the index of real agricultural production (a detailed analysis of the characteristics of the series is given in [29]). Data from 2000–2020 are used for training. The series has annual seasonality. The test period is 2021. The results are summarized in Table 2. The best models in terms of forecast accuracy for the test period or the value of the logarithm of the maximum likelihood function are highlighted in bold font.

The lowest values for the logarithm of the maximum likelihood function have the hybrid model with trend estimation by the ACD algorithm with information on the seasonality of the time series and its prediction for the test period using ETS. However, this method does not have the most accurate results of forecasts for the test period (MAE and RMSE columns).

The best MAE results for forecasting for the test period are obtained using the method of trend estimation

Table 2. Models of the index of real volume of agricultural production according to macroeconomic statistics of the Russian Federation and their forecasts for the test period

Trend estimation	LLF	MAE	RMSE
1st degree polynomial $y = a_0 + a_1x$	−395.597	0.084	0.097
2nd degree polynomial $y = a_0 + a_1x + a_2x^2$	−395.871	0.096	0.105
3rd degree polynomial $y = a_0 + \sum_{i=1}^{3} a_i x^i$	−396.132	0.112	0.117
4th degree polynomial $y = a_0 + \sum_{i=1}^{4} a_i x^i$	−396.173	0.114	0.119
5th degree polynomial $y = a_0 + \sum_{i=1}^{5} a_i x^i$	−397.175	0.171	0.185
1st degree polynomial + ARIMA	−395.567	0.083	0.096
2nd degree polynomial + ARIMA	−395.734	0.090	0.100
3rd degree polynomial + ARIMA	−395.830	0.093	0.103
4th degree polynomial + ARIMA	−395.836	0.094	0.103
5th degree polynomial + ARIMA	−396.031	0.102	0.107
Hyperbolic function $y = a_0 + a_{-1} / x$	−394.600	0.079	0.124
Logarithmic function $y = a_0 + a_1\ln x$	**−395.058**	**0.069**	**0.101**
Exponential function $y = \exp(a_0 + a_1x)$	**−395.310**	**0.075**	**0.095**
Interpolation by splines	−396.028	–	–
Fourier series transformations	**−395.648**	**0.076**	**0.089**
Exponential smoothing	−395.343	0.078	0.096
LOESS method	−395.379	0.079	0.097
ACD algorithm	−396.643	0.145	0.153
ACD + ARIMA trend	−397.140	0.170	0.195
ACD + ETS trend	−400.076	0.121	0.121

Table 3. Time series models of RF Sberbank stocks and their forecasts for the test period

Trend estimation	LLF	MAE	RMSE
1st degree polynomial $y = a_0 + a_1x$	−4979.271	0.105	0.128
2nd degree polynomial $y = a_0 + a_1x + a_2x^2$	−4950.327	0.199	0.215
3rd degree polynomial $y = a_0 + \sum_{i=1}^{3} a_i x^i$	−4881.779	0.440	0.465
4th degree polynomial $y = a_0 + \sum_{i=1}^{4} a_i x^i$	−4986.211	0.108	0.114
5th degree polynomial $y = a_0 + \sum_{i=1}^{5} a_i x^i$	−5019.458	0.079	0.119
1st degree polynomial + ARIMA	−4993.657	0.070	0.093
2nd degree polynomial + ARIMA	-4995.075	0.067	0.089
3rd degree polynomial + ARIMA	−4997.347	0.063	0.083
4th degree polynomial + ARIMA	−5006.537	0.067	0.085
5th degree polynomial + ARIMA	−5006.588	0.062	0.079
Hyperbolic function $y = a_0 + a_{-1} / x$	−4880.421	0.443	0.451
Logarithmic function $y = a_0 + a_1 \ln x$	−4923.877	0.288	0.300
Exponential function $y = \exp(a_0 + a_1x)$	−4992.214	0.072	0.094
Interpolation by splines	−5013.558	–	–
Fourier series transformations	−5014.470	0.038	0.054
Exponential smoothing	−4976.059	0.117	0.145
LOESS method	**−5013.578**	**0.004**	**0.006**
ACD algorithm	−4996.171	0.077	0.086
ACD + ARIMA trend	−4987.216	0.145	0.162
ACD + ETS trend	−4979.289	0.170	0.191

using logarithmic and exponential functions (the best models are highlighted in bold). Models with Fourier series expansion, exponential smoothing, hyperbolic function, as well as those based on the LOESS method, have close values to them.

The best results in terms of RMSE for forecasting for the test period are those obtained using the method of trend estimation based on Fourier series expansion. The error values of trend modeling with exponential function, exponential smoothing, LOESS method and with polynomial of the first degree with forecasting both on the basis of ARIMA and linear regression are acceptable.

The lowest value of the likelihood function after the hybrid model with ACD trend and ETS forecast is obtained using the method of modeling the trend based on a polynomial of the fifth degree with linear regression forecasting. This value is slightly higher for the hybrid model with ACD trend and time series seasonality information predicted by ARIMA. We can also note the models with trend estimation using third- and fourth-degree polynomials. However, this group of models has some of the worst indicators for forecast accuracy.

When estimating the trend using polynomials, it is important to note that the accuracy decreases as the degree of the polynomial increases. This property also holds when combining trend estimation using polynomials with the ARIMA model.

In the examples presented above, seasonal time series are processed. Let us consider the performance of various methods of trend estimation on a non-seasonal time series of RF Sberbank stocks[4] (rubles). Data from 2000–2021 are used for training. The test period is 2022. Graphs of ACF and PACF functions are presented and described in detail in [29, Fig. 7]. The modeling and forecasting results are presented in Table 3. The best

[4] Sberbank (SBER) stock price. https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER (in Russ.). Accessed November 03, 2023.

models in terms of prediction accuracy for the test period or the value of the logarithm of the maximum likelihood function are highlighted in bold font.

Since there is no seasonality for stock market data, the forecast result depends entirely on the trend model and the random component.

The behavior of the series is described best by the spline function and LOESS models. At good values of the likelihood function, the trend model with the LOESS method has the most accurate forecasts. The error is 0.004 for MAE and 0.006 for RMSE. Low LLF values also correspond to trend modeling using a fifth-degree polynomial with linear regression prediction and using Fourier series expansion. These models also have some of the best accuracy values after the LOESS method model.

In contrast to the modeling of seasonal series, the other models are less well adjusted to the behavior of the series. These models are characterized by deterioration of accuracy as LLF values increase.

In trend modeling with polynomial, as the degree of the polynomial increases, starting from the fourth degree, accuracy increases, which may be partly due to overtraining. This is also true for the combination of series modeling with polynomial and ARIMA; however, the accuracy performance improves as the degree of the polynomial increases from the first degree onwards.

CONCLUSIONS

When modeling the trend for seasonal series, overlaying information on seasonal and random components affects the quality of forecasts. The best results are shown by the methods of trend modeling using Fourier series expansion and LOESS method. The combination of trend modeling with polynomial and ARIMA method for seasonality also has quite accurate results. While the indicators are worse when using polynomials for trend estimation with linear regression forecast than when using the combination of polynomial with ARIMA model, the behavior dynamics of accuracy indicators is the same for them.

It is interesting to note that the ACD algorithm performed best on the data of monetary incomes of the population. This time series has heterogeneous dispersion. Data forecasting using the ACD algorithm can be very useful for heteroscedastic series.

Trend modeling using exponential and logarithmic functions did not demonstrate outstanding results. Such methods are also computationally more complex than their polynomial equivalents. The logarithmic function model has limitations on the data values due to the lack of a real logarithm of a negative argument. The hyperbolic function model is one of the worst performing, both in terms of likelihood function and accuracy estimates.

Unlike trend estimation using the exponential function, trend extraction using exponential smoothing led to one of the best results. However, the main disadvantage of this method is the uncertain smoothing coefficient.

When working with non-seasonal time series, the forecast quality depends only on the trend and noise components. Models with Fourier series expansion, LOESS method and spline function fit the data best. However, it is difficult to make forecasts based on splines due to their orientation to data interpolation.

Since the LOESS method for a group of non-seasonal series also gives the best or close to the best results for all indicators, this method can be recommended for obtaining the most accurate results for series of different nature. The modeling of the trend using Fourier series decomposition can also be highlighted based on the sufficiently accurate results for time series of different natures obtained using this approach.

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

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
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., Petrusevich 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. Petrusevich 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 *βSARMA* 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. Gramovich I.V., Musatov D.Yu., Petrusevich 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

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. Petrusevich, 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

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

Петрусевич 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

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt