МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ В ПРИРОДНИЧИХ НАУКАХ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

DOI: 10.31319/2519-8106.2(41)2019.185017
UDC 519.688
A.O. Dolhikh, postgraduate student, dolgih.nastya@gmail.com
O.G. Baibuz, doctor of Engineering Science, Prof., obaybuz@ua.fm
Dniprovsky National University named after Oles Honchar, the city of Dnipro

THE SOFTWARE DEVELOPMENT FOR TIME SERIES FORECASTING WITH USING ADAPTIVE METHODS AND ANALYSIS OF THEIR EFFICIENCY

Searching algorithm for optimal values of the smoothing coefficients of adaptive models of time series forecasting by the genetic algorithm is described. The results of the proposed approach in forecasting financial indicators are presented. The analysis of the effectiveness results of the developed algorithm with the help of a multi-criterion procedure is carried out, which allows to consider the accuracy of forecasts, the complexity of the model and to conduct analysis of adequacy using Fisher test, the determination coefficient and the mechanism for checking the residues.
Keywords: *forecasting; adaptive methods; genetic algorithm; analysis of model quality.*

Описаний алгоритм пошуку оптимальних значень коефіцієнтів згладжування адаптивних моделей прогнозування часових рядів за допомогою генетичного алгоритму. Представлені результати роботи запропонованого підходу у прогнозуванні фінансових показників. Виконаний аналіз результатів ефективності розробленого алгоритму за допомогою багатокритеріальної процедури, яка дозволяє враховувати точність прогнозів, складність моделі та проводити аналіз адекватності з використанням тесту Фішера, коефіцієнту детермінації та механізму перевірки залишків.
Ключові слова: *прогнозування; адаптивні методи; генетичний алгоритм; аналіз якості моделі.*

Setting of a problem

The task of forecasting is particularly relevant in various areas of human activity. In the economy — to predict daily fluctuations in stock prices, exchange rates, weekly and monthly sales volumes, annual production volumes, etc. In natural sciences — to predict the amount of rainfall, natural phenomena, pollution of water resources, the assessment of some biological and biochemical indicators.

Forecasting with many known methods requires substantial mathematical calculations and imposes limitations on the size of a series. Other methods are devoid of these shortcomings, but do not make it possible to perform a fairly accurate forecast. In addition, the difficulty in developing reliable forecasting models can be related to the complex nature of the relevant statistics, if they are characterized by a sharp change in the dynamics of the indicator and non-linearity. In such conditions, forecasting models based on the principles of exponential smoothing are promising. They have a number of advantages over other statistical models: easy to build on experimental data; their application does not require high costs of machine time and complex mathematical calculations; they take into account the "aging" of information. However, in applying this approach, it is important to find the optimal values of smoothing coefficients $в_1$, $в_2$, $в_3$ for a given series. Therefore, the issue of software development for automatic selection of these coefficients and comparison of the obtained results with the results of other statistical forecasting models is relevant.

Analysis of recent research and publications

According to the current estimates [1, 2], for today there are about one hundred types of models of time series forecasting. Researchers are constantly improving existing methods of forecasting and developing new ones. In connection with this, in the works [1, 3] it is proposed to classify them into the following groups: model of the subject area, that is, those which use certain laws of physics, biology, etc. for the construction of the forecast, and models of time series, which are universal for various subject areas. In turn, time series models are also divided into two groups: statistical and structural. Statistical models are considered those ones in which the dependence of the future value from the past is given in the form of some equation [1]. These include regression and auto-regression models, exponential smoothing models, and so on. Structural models are those which try to find some regularities in the development of the process within the time series [3], for example, models based on fuzzy logic, decision trees, neural networks, Markov chains, reference vectors, etc.

The current study examines the issue of forecasting financial indicators using statistical models. The most well-known forecasting models from this class are regression models. In such models, a series is presented as a sum or product of three components:

$$\left[\begin{array}{l} f(t) = T(t) + S(t) + A(t) + \xi_t \\ f(t) = T(t) * S(t) * A(t) * \xi_t \end{array}\right. \qquad (1)$$

where *T (t)* is a trend line that shows the global changes in the phenomenon under study, *S (t)* is the seasonality that reflects fluctuations relative to the trend due to external influences; *A (t)* — cyclicity (auto-oscillations) — more or less regular fluctuations relative to the trend, due to the internal nature of the phenomenon under study; o_t — forecast error. Forecasting of time series in the case of regression models is to identify the forms *T* (*t*), *S* (*t)* and *A* (*t*). The essential disadvantage of these models is the need for complex mathematical calculations [4]. In addition, to construct correct models, it is important to have a large database of observations and to correctly select the shape of the regression line [4].

Today methods which represent further development of the method of regression analysis, in particular GMDH — group method of data handling are gaining popularity. In the prediction of time series using GMDH, the process is usually presented in the form of 2:

$$y = a + \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} x_i x_j , \qquad (2)$$

where *m* is the number of variables, a, b_i, c_{ij} are coefficients of variables in polynomials, also called weights, *y* is the value of the series for which forecasting is being built, x_i and x_j are the previous values of the time series [5]. During the training, optimum parameters are found. In [5,6] the construction of neural networks of the GMDH type is considered. The weights of the neurons of the constructed network are used as coefficients models. Neural networks of the GMDH type are implemented in the GMDH library of a well-known statistical data processing package R [5]. The main problem while using this approach is, this method like all regression models, requires a rather large representative sample and complex mathematical calculations for a qualitative result [7].

Auto-regressive methods make it possible to get rid of the need to have a fairly large learning sequence and to choose the optimal representation of regression forms. Today there is a hierarchy of methods of auto-regression-running mean, which can be logically defined as follows [8]:

$$AR(p) + MA(q) \to ARMA(p,q) \to ARIMA(p,q,d), \qquad (3)$$

where the *AR* (*p*) is the autoregressive model of the order *p, MA* (*q*) is the model of the moving average-range order *q, ARMA* (*p, q*) is the autoregressive model — of moving average with the order of the auto-regression *p* and the order of the moving average *q, ARIMA* (*p, q, d*) — integrated model *ARMA* (*p, q*) with integration order d. Models *ARIMA* (*p, q, d*) are the most well-known models in the auto-regressive class. When used in forecasting, a time series is presented in the form of 4:

$$\Delta^d X_t = c + \varepsilon_t + \sum_{i=1}^{p} \alpha_i \Delta^d X_{t-i} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}\,, \tag{4}$$

where *c, бi, bj* are parameters of the model, *Дd* — operator of the time series difference order *d, p* — autoregressive order, *q* — the order of the moving average, e_t — independent, normally distributed errors with zero average. Models *ARIMA* (*p, q, d*) make it possible to switch from non-stationary to stationary process. The advantages of ARIMA forecasting models are transparency of modeling; uniformity of analysis and design and variety of applications. The disadvantages are labor inputs and resource intensity of identifying the most appropriate model and the impossibility of modeling nonlinear dependencies. In [9], an algorithm for identifying the optimal form of the ARIMA model is proposed by sorting the values of the order of auto-regression, moving average and integration.

In contrast to the above listed statistical methods of forecasting, adaptive models neither impose constraints on the size of a series, nor require the execution of complex mathematical operations, enable the modeling of not only linear dependencies. In addition, such models are based on the principles of exponential smoothing, that is they take into account the "aging" of information. In this approach, a series is presented as a function:

$$u_t = f(a_{1t}, a_{2t}, ..., a_{pt}, t) + e_t\,, \tag{5}$$

where *t* — time indicator; a_{1t}, a_{2t}, ..., a_{pt} — coefficients of the adaptive model at the moment of time *t* [10]. Eight types of adaptive forecasting are distinguished, depending on the form of the trend and the presence or absence of a periodic component. In this approach, the question of finding the optimal values of the smoothing parameters $в_1$, $в_2$, $в_3$ is used to calculate the coefficients a_{1t}, a_{2t}, ..., a_{pt}. When the values of $в_1$, $в_2$, $в_3$ change, the forecasting error value can increase significantly. In conditions of uncertainty, some authors suggest to accept $в_1 = в_2 = в_3 = 0.3$ [4]. But in such an approach one cannot achieve the required quality of forecasting. Therefore, it is promising to identify the optimum values of the smoothing parameters. In the current study, an algorithm for solving this problem is proposed.

In addition, it is also important to analyze the effectiveness of the adaptive approach and other method-representatives of the statistical class, which are more time-consuming. To do this, the authors developed a multi-criteria procedure that allows considering the accuracy of forecasts, the complexity of the model and its adequacy and compliance with the process under study.

Objectives setting

The purpose of the research is to develop software for identifying the optimal values of the smoothing coefficients of adaptive methods for forecasting time series. For a thorough analysis of the importance of the results obtained, it is worth comparing the results of the proposed algorithm with the results of other statistical models. To achieve the goal, the following tasks were set:

1. To develop an algorithm for finding optimal values of parameters $в_1$, $в_2$, $в_3$.
2. To apply the proposed development to forecast financial indicators.
3. To compare the results of forecasting, obtained by the adaptive methods without using searching algorithm for finding optimal values, proposed by the authors and with its usage.
4. To develop a multi-criteria procedure for estimating the effectiveness of forecasting models, which allow taking into account the accuracy of forecasts, the complexity of the model and its adequacy and compliance with the investigated process. Compare the results of the implemented algorithm with the results of the work of other regression methods of forecasting by means of the developed procedure.

Presentation of the baseline

In adaptive models, a time series is presented as a function 5, during the time of which the value of the deviations of predictive values from the values of the initial series is monitored [10]. They are divided into two groups: linear and seasonal. In models of linear growth, the forecast is calculated by the formula:

$$u_{t+\tau} = a_{1t} + a_{2t}\tau\,, \tag{6}$$

where *a* — the number of steps of the forecast; a_{1t}, a_{2t} — the coefficients of the adaptive model at a moment of time *t*. The adaptive models of linear growth include the Holt model, the Tail-Wage model, Brown's model, and the Box-Jenkins model. Seasonal adaptive models, besides the trend, allow taking into account periodic fluctuations. Linear adaptive, linear multiplicative (Winters), exponential additive and exponential multiplicative are among them. The listed models differ in the means of finding the parameters a_{1t}, a_{2t}. For example, for the Holt model, formulas 7 are used, and for the model Tail-Wage — formulas 8.

$$\begin{cases} a_{1,t} = \beta_1 u_t + (1-\beta_1)(a_{1,t-1} + a_{2,t-1}) \\ a_{2,t} = \beta_2 (a_{1,t} - a_{1,t-1}) + (1-\beta_2) a_{2,t-1} \end{cases} \quad (7)$$

$$\begin{cases} a_{1,t} = \beta_1 u_{t-1} + (1-\beta_1)\hat{u}_t \\ a_{2,t} = a_{2,t-1} + \beta_1 \beta_2 e_t \\ e_t = u_t - \hat{u}_t \end{cases} \quad (8)$$

where $в_1$, $в_2$, $в_3$ are the smoothing coefficients that take values from 0 to 1, u_t — the real value of the series level at t-th step, u_t — the predictive value at t-th step, e_t — the error at the t-th step.

A characteristic feature of all adaptive models is that at each step new values of a_{1t}, a_{2t} are calculated. In order to obtain qualitative forecasting results, it is necessary to find the values of the parameters $в_1$, $в_2$, $в_3$, which most correspond to the given series. To solve this problem software was implemented that would allow finding the optimal values of $в_1$, $в_2$, $в_3$ by means of a genetic algorithm.

The potential solution of the optimization problem (parameters $в_1$, $в_2$, $в_3$) is presented as chromosomes. In classical genetic algorithms, the chromosomes are presented in the form of a binary vector. However, the binary representation of the chromosomes entails certain difficulties in the search of continuous spaces, which are associated with the large dimensionality of the search space [11]. Therefore, a genetic algorithm with encoding based on real numbers (RGA) was used to solve this problem. Below is the algorithm proposed by the authors for forecasting the time series using the genetic algorithm and adaptive methods.

Forecasting of time series using adaptive methods and genetic algorithm

Step 1: At the first step, the initial population of the chromosomes is generated. Population size, K, is given by the user. Each chromosome is a set of values $в_1$, $в_2$, $в_3$.

Step 2: For each chromosome in a population, mean square error (MSE) of forecast using the formula is calculated:

$$MSE = \frac{\sum_{t=1}^{n} (ut - \hat{ut})^2}{n - m - 1}, \quad (9)$$

where n — the length of a series, m — the number of model parameters. To do this, using the parameters $в1$, $в2$, $в3$, based on the adaptive model selected by the user, a forecasting series is build. Hereupon, one calculates the mean square error of forecast, *MSE*. The smaller the *MSE* value, the better the model is.

Step 3: Select two chromosomes from the population according to the strategy given by the user and perform chromosomes crossing. As a result of this operation, there are two new individuals who inherit the signs of the parents. Both "parents" and both "descendants" are added to the "intermediate" population. Parents' selection and crossing are conducted M times. M is set by the user.

Step 4: Carry out a mutation. For each chromosome from the population, a random number is generated. If it is less than the mutation level, then chromosome mutation is carried out according to one of the possible strategies: simple mutation or Michalewicz mutation. The strategy can be changed using the program interface. By default, a simple mutation is used.

Step 5: At this step, selection or sampling of the chromosomes for the next generation is carried out. For the obtained "intermediate" population, one calculates the *MSE* value. A certain amount of chromosomes with the lowest value of the mean square error forecast, q, enter the next population. In order to prevent early degeneration of the population, random K-q chromosomes are generated and added to the new population.

Steps 2—5 are performed iteratively until one of the conditions for stopping the genetic algorithm is fulfilled: the number of generations entered by the user is reached or the algorithm will come up to a single solution. After that the set of values of coefficients ($в_1$, $в_2$, $в_3$) is used for prediction of the following values of a series.

Testing of the of the developed results software was carried out in time series, which represent time series of shares of well-known foreign companies such as Apple, Microsoft, IBM, etc., at the close of exchange from December 2017 till December 2018 [12]. Fig. 1 shows the time series of price of Microsoft and its correlogram, which suggests that there is a growing trend in the series and there are no pronounced cyclical fluctuations.

a) b)

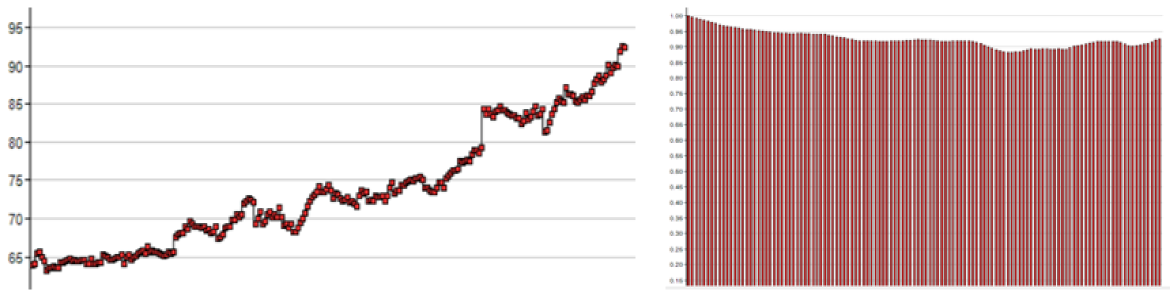

Fig. 1. Series (Microsoft): a) — output value of series, b) — correlogram of series

In conditions of uncertainty, it is recommended to use $в_1 = в_2 = в_3 = 0.3$ as values of coefficients [4]. Tabl. 1 shows the mean square error forecast, *MSE* for predicting Microsoft series using values of default coefficient (0.3) and *MSE* values if used the approach proposed by the authors. For the adjustment of models, a genetic algorithm with valid coding by the following parameters was used: N = 100 (population size), percentage of the best chromosomes to be left in population, q = 80%, number of epochs (repetitions) of the genetic algorithm = 30, mutation level = 0, 1 Table 2 shows how much the *MSE* value decreased when using the approach proposed by the authors compared to the recommended default values.

Table 1. Mean square error forecast, *MSE*

Model	***MSE***	***Model***	***MSE***
Holt model ($в_1$=$в_2$=$в_3$=0.3)	1.0232	Holt model ($в_1$=0.9156, $в_2$=0.0552)	0.581
Tail-Wage model ($в_1$=$в_2$=$в_3$=0.3)	1.0232	Tail-Wage model ($в_1$=0.9142, $в_2$=0.0564)	0.6002
Brown model ($в_1$=$в_2$=$в_3$=0.3)	1.0197	Brown model ($в_1$=0.7112)	0.8195
Box-Jenkins model ($в_1$=$в_2$=$в_3$=0.3)	0.8353	Box-Jenkins model ($в_1$=0.8327, $в_2$=0.0742,$в_3$=0.0654)	0.6185
Linear additive ($в_1$=$в_2$=$в_3$=0.3)	0.7093	Linear additive ($в_1$=0.874, $в_2$=0.0087,$в_3$=0.0397)	0.6266
Linear multiplicative ($в_1$=$в_2$=$в_3$=0.3)	0.7532	Linear multiplicative ($в_1$=0.00, $в_2$=0.0071,$в_3$=0.8814)	0.6227
Exponential additive ($в_1$=$в_2$=$в_3$=0.3)	0.7201	Exponential additive ($в_1$=0.0509, $в_2$=0.0912,$в_3$=0.8773)	0.5854
Exponential multiplicative ($в_1$=$в_2$=$в_3$=0.3)	0.7204	Exponential multiplicative ($в_1$=0.2005, $в_2$=0.0147,$в_3$=0.8543)	0.5939

Table 2. Decrease of value of mean square error forecast, *MSE*

Model	***Decrease of value of mean square error forecast, %***
Holt model	43.22
Tail-Wage model	41.34
Brown model	19.63
Box-Jenkins model	25.98
Linear additive	11.66
Linear multiplicative	17.33
Exponential additive	18.71
Exponential multiplicative	17.56

As shown in tabl. 1 and 2, using the proposed approach, it is possible to achieve a reduction in the mean-square error forecast at least by 11.66%. In the best case (for the Holt model) we managed to improve the quality of the built forecasts by 43.22%.

However, in order to fully assess the significance of the results obtained, it is worth comparing the results of the adaptive models with the results of other models of time series, which are also representatives of the class of regression methods and have proven themselves well in the analysis and forecasting of processes different by nature and complexity. For this purpose, integrated models of autoregression integrated moving average [9] (ARIMA) and group method of data handling (GMDH) [5,6] were used. These models were previously trained in the same time sequences as adaptive methods. The optimal values of the integration order, autoregression, and moving average of ARIMA model were found using procedures [9,13]. The training of the neural network of the GMDH type was conducted using the algorithm [5].

In order to compare the quality of forecasting, a multi-criteria procedure was developed that allows to regard the accuracy of built forecasts, the complexity of the model, and the results of the analysis of the residues. Forecasting accuracy is estimated by calculating the mean square forecast error, *MSE*, and sum squared error, *SSE.* To assess the relationship between the accuracy of the model and its complexity — the significance of the information quality criteria of Akaike, AIC, and Schwartz, *BIC.* The smaller the value of *SSE, MSE, AIC* and *BIC*, the better is the result. The adequacy of the model is verified using Fisher test and the calculation of the value of the adjusted determination coefficient, R^2*. The higher the value of the determination coefficient, the more qualitative is the model. Tabl. 3 shows the results of comparing the forecasting efficiency using adaptive models, ARIMA models, and group argumentation based on *SSE, MSE, AIC, BIC* and R^2*.

Table 3. Results of model quality comparison by quantitative criteria

Model	***SSE***	***MSE***	***AIC***	***BIC***	R^{2*}	***Number of parameters***
GMDH	146.4159	0.9893	0.2831	1.7495	98.0949	105
Brown model	205.6839	0.8195	-0.1912	-0.1633	97.329	1
Lin. ad.	154.4175	0.6177	-0.47	-0.4281	97.9913	3
Lin. mult.	153.1447	0.6126	-0.4783	-0.4364	98.0078	3
Exp. ad.	146.1417	0.5846	-0.5251	-0.4832	98.0985	3
Exp. mult.	146.1336	0.5845	-0.5252	-0.4833	98.0986	3
Box-Jenkins model	145.3183	0.5813	-0.5307	-0.4886	98.1092	3
Tail-Wage model	145.527	0.5798	-0.5372	-0.5093	98.1064	2
Holt model	145.5264	0.5798	-0.5372	-0.5093	98.1064	2
ARIMA	131.1438	0.5309	-0.6097	-0.5259	98.2928	6

By comparing the effectiveness of the implemented forecasting models shown in table 3, one can conclude that for Microsoft series the best results have been shown by the ARIMA. But among the class of adaptive models, one can also find the model that will give sufficiently high-quality forecastings. For example, if you compare the values of *MSE* and R^2 * obtained by the ARIMA model and the Holt model in the percentage scale, you can easily see that the *MSE* value for the Holt model is greater than the *MSE* for the ARIMA model by 9.22%, and the value of the determination coefficient for the ARIMA model is higher than R^2 * for the Holt model only by 0.2%. An important factor here is that the use of the ARIMA model requires the calculation of the values of six parameters, while the use of the Holt model requires the calculation of only two coefficients. It is worth noting that all adaptive models by the values of information quality criteria, *AIC* and *BIC*, show better results than models of neural networks of type GMDH, which is caused by too many parameters of GMDH.

Fig. 2 shows the time series that represents the stock prices of AAON Inc. in the period from 2017 to 2018 and its correlogram. Looking at them one can draw a conclusion about a distinct seasonal component in a series and a lack of a tendency to increase or decrease. Tabl. 4 presents the results of the evaluation of the effectiveness of statistical models implemented in software, conducted with the help of a multi-criteria procedure.

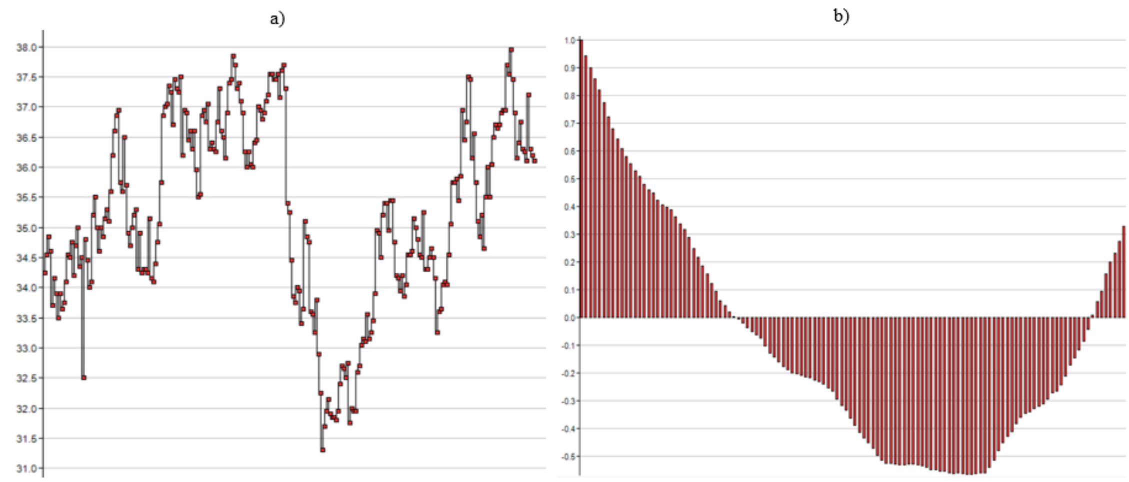

Fig. 2. Series (AAON Inc.): — output value of a series, b) — correlogram of a series

Table 4. Results of model quality comparison by quantitative criteria

Model	***SSE***	***MSE***	***AIC***	***BIC***	R^{2*}	***Number of parameters***
GMDH	64.4917	0.4425	-0.5214	0.945	80	105
Brown model	100.682	0.4011	-0.9056	-0.8777	70.1742	1
Exp. ad.	72.7227	0.2909	-1.223	-1.1811	77.9284	3
Exp. mult.	72.7128	0.2909	-1.2232	-1.1813	77.9313	3
Box-Jenkins model	72.0692	0.2883	-1.232	-1.1901	78.1146	3
Tail-Wage model	72.0387	0.287	-1.2404	-1.2124	78.1233	2
Holt model	72.0365	0.287	-1.2404	-1.2125	78.1239	2
ARIMA	69.8501	0.282	-1.2475	-1.1777	78.7482	5
Lin. mult.	69.0402	0.2762	-1.275	-1.2331	78.98	3
Lin. ad.	69	0.276	-1.2756	-1.2337	78.9916	3

As one can see from tabl. 4, for the time series with a pronounced periodic component, according to the values of information quality criteria, adaptive models have shown significantly better results than more complicated statistical methods. Only by the value of the *SSE* they yield to the models of group model of data handling. However, the use of GMDH requires the calculation of a much larger number of parameters (105 and 3), while it gives reduction of the error value only by 6.5% as compared to the linear additive model.

Conclusions and perspectives of further research

1. In the course of the research there was developed the software for identifying the optimum values of smoothing coefficients of adaptive forecasting methods, $в_1$, $в_2$, $в_3$. For this purpose, a genetic algorithm with valid coding was used.

2. The proposed approach was used to forecast time series that represent the daily fluctuations of stock prices of known American companies such as IBM, Microsoft, Apple, AAON, etc., from December 2017 till December 2018. Using searching algorithm of optimal values of the smoothing coefficients, it was possible to improve the accuracy of forecasts not less than 10%.

3. The developed software complex allows comparing the results of various time series forecasting models using the multi-criteria procedure, which makes it possible to take into account error of forecasting, *SSE* and *MSE*, information quality criteria, *AIC* and *BIC*, and adjusted determination criterion, R^2*. Performance analysis of adaptive methods was compared with the results of the ARIMA and GMDH models. The analysis of the results showed that in the most series by the sum square error forecast they are inferior to the models of autoregression integrated moving average and group method of data handling. However, if we take into account not only the value of error but also the complexity of mathematical calculations, then adaptive methods in all cases show better results than group method of data handling that uses too many parameters, and in certain ones they exceed ARIMA models, for example in the analysis of sequences with a pronounced seasonality. This also confirms the importance and complexity of the question of selecting the optimal forecasting model for a specific time series [14].

For today, the perspective direction of research in the field of time series analysis is the question of construction of structural models such as neural networks, fuzzy logic, Markov chains, etc. and analysis of their efficiency. Therefore, the purpose of the further research is to expand the developed software complexwith forecasting models of this class, as well as to compare the quality of the results of statistical and structural models by means of a multi-criteria procedure. Such an approach will enable to perform the automatic selection of optimal forecasting models for a specific dynamic series. In addition, the best models selected can be used later in the construction of ensembles, that is, sets, models of forecasting, which can significantly improve the forecasting quality [15].

References

[1] Chuchuieva, I.A. (2012). Model' prohnozuvannya chasovykh ryadiv za vybirkoyu maksymal'noyi pravdopodoby: dysertatsiya na otrymannya stupenyu kandydata tekhnichnykh nauk [A time series forecasting model for a maximum likelihood sample: thesis for a candidate of technical sciences]. *Candidate's thesis*. Moscow: Moscow State Technical University named after Bauman.

[2] Tikhonov, E.E. (2006). Metody prohnozuvannya v umovakh rynku [Forecasting methods in market conditions]. Nevinnomyssk.

[3] Yang, J.M. (2006). Power System Short-term Load Forecasting. *Candidate's thesis*. Germany, Darmstadt: Elektrotechnik und Informationstechnik der Technischen Universitat.

[4] Biloborodko, O.I., T.G. Yemelyanenko. (2014). Analiz dynamichnykh ryadiv [Dynamic Series Analysis]. Dnipro: DNU named after Oles Holnchar [in Ukrainian].

[5] Osman Dag, Ceylan Yozgatligil. (2016). GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms. *The R Journal Vol. 8/1, Aug. 2016, ISSN 2073-4859*. Retrieved from https://journal.r-project.org/archive/2016/RJ-2016-028/RJ-2016-028.pdf.

[6] Dolhikh A.O. (2019). Prohnozuvannya finansovykh chasovykh ryadiv z vykorystannyam neyronnoyi merezhi typu MHUA [Financial time series forecasting using a neural network of the GMDH type]. Proceedings from *XXIII mizhnarodnoho molodizhnoho forumu «Radiolektreronika i molod□ u XXI stolitti», konferentsiya «Matematychne modelyuvannya, systemnyy analiz i teoriya optymal'nykh rishen'» – The Twenty Third International Youth Forum "Radio Electron-*

ics and Youth in the XXI Century", Conference "Mathematical Modeling, Systems Analysis and Theory of Optimal Solutions". (pp. 146–147). Kharkiv [in Ukranian].
[7] The group method of data handling (GMDH). Retrieved from http://www.mgua.irtc.org.ua/ru/index.php?page=gmdh (application date: 09.05.2019)
[8] A.V.Tregub, I.V.Tregub. (2011). Metodyka pobudovy modeli ARIMA dlya prohnozuvannya dynamiky chasovykh ryadiv. [The method of constructing the ARIMA model for predicting the dynamics of time series]. «Lisovyy visnyk», T.5, rozdil «Matematychne modelyuvannya» – *"Forecast Newspaper", 5, section "Mathematical modeling"*. (pp. 179–183).
[9] Hyndman, R.J., Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software. July 2008. Vol. 27, Issue 3*.
[10] Lukashin, Yu.P. (2003). Adaptyvni metody korotkostrokovoho prohnozuvannya chasovykh ryadiv [Adaptive methods of time series short-term forecasting]. M.: Finance and Statistics.
[11] Herrera F., Lozano M., Verdegay J.L. (1998). Tackling real-coded Genetic algorithms: operators and tools for the behavior analysis. *Artificial Intelligence Review, Vol. 12, No. 4,* (pp. 265–319).
[12] YAHOO! Finance. Retrieved from https://finance.yahoo.com/.
[13] Kwiatkowski, D., Phillips P.C.B., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*. (pp. 159–178).
[14] Armstrong, J.S., Adya, M., Collopy, F. (2001). Rule-Based Forecasting Using Judgment in Time Series Extrapolations. Principles of Forecasting: A Handbook for Researcher and Practitioners. Kluwer, Dordrecht.
[15] Wichard, J.D., Ogorzalek, M. (2007). Time series prediction with ensemble models. *Neurocomputing, 70, Issues 13–15. (pp. 2371–2378).*

РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ З ВИКОРИСТАННЯМ АДАПТИВНИХ МЕТОДІВ ТА АНАЛІЗ ЇХ ЕФЕКТИВНОСТІ

Долгіх А.О. Байбуз О.Г.

Реферат

Особливої актуальності в різних областях людської діяльності набуває задача прогнозування. В економіці — для прогнозування щоденних коливань цін на акції та курсів валют. В природничих науках — для визначення кількості опадів, оцінки біологічних та біохімічних показників.

Прогнозування за допомогою багатьох відомих методів вимагає суттєвих математичних обчислень та накладає обмеження на розміри ряду. Інші методи позбавлені цих недоліків, але не дозволяють робити достатньо точний прогноз. В таких умовах перспективним є використання адаптивних моделей. Їх легко будувати на експериментальних даних; їх застосування не потребує великих витрат машинного часу та складних математичних обчислень; вони враховують «старіння» інформації. Однак при застосуванні цього підходу важливо віднайти оптимальні значення коефіцієнтів згладжування, $в_1,в_2,в_3$ для заданого ряду. Актуальним є питання розробки програмного забезпечення для автоматичного підбору цих коефіцієнтів.

Метою дослідження є розробка програмного забезпечення ідентифікації оптимальних значень коефіцієнтів згладжування адаптивних методів прогнозування та оцінка якості отриманих результатів. Для ідентифікації оптимальних значень $в_1,в_2,в_3$, була розроблена процедура пошуку, заснована на використанні генетичного алгоритм з дійсним кодуванням. Запропонований підхід був використаний для прогнозування щоденних коливань цін на акції відомих компаній, таких як IBM, Microsoft, Apple, AAON, тощо. Для порівняння ефективності адаптивних методів з результатами роботами інших статистичних підходів була багатокритеріальна процедура. Вона дозволяє враховувати значення похибок прогнозу, інформаційних критеріїв якості, й

скорегованого критерію детермінації. Результати порівняння показали, що, при використанні невеликої кількості параметрів у порівнянні з іншими моделями, адаптивні методи можуть будувати досить якісні прогнози. Отримані результати можуть бути у подальшому використані для ідентифікації оптимальної моделі динамічного процесу та при побудови ансамблів моделей прогнозування.

Література

1. Чучуєва І. А. Модель прогнозування часових рядів за вибіркою максимальної правдоподоби: дис. … канд. техн. наук: 05.13.00 / Московський державний технічний університ. Москва, 2012. 154 с.
2. Тихонов Е. Є. Методи прогнозування в умовах ринку: учбовий посібник .Невинномиськ, Північно-Кавказький державний технічний університет, 2006. 221 с.
3. Jingfei Y. M. Power System Short-term Load Forecasting: Candidate dissertation / Technical University Of Electrical and Computer Engineering. Darmstadt, Germany, 2006. 139 p.
4. Білобородько О. І., Ємел'яненко Т. Г. Аналіз динамічних рядів: учбовий посібник. Дніпро: РВВ ДНУ, 2014. 80 с.
5. Dag O., Yozgatligil C. GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms. *The R Journal*. 2016. Vol. 8. No. 1. P. 379-386.
6. Долгіх А. О. Прогнозування фінансових часових рядів з використанням нейронної мережі типу МГУА: *Сучасні методи обробки зображень*: матеріали XXIII міжнародного молодіжного форуму «Радіоелектроніка і молодь у XXI столітті». Харків: ХНУРЕ, 2019. Т. 7. С. 146–147.
7. Метод групового урахування аргументів (МГУА). URL: http://www.mgua.irtc.org.ua/ru/index.php?page=gmdh (дата звернення: 11.10.2019).
8. Трегуб А. В., Трегуб І. В. Методика побудови моделі ARIMA для прогнозування динаміки часових рядів. *Лісовий вісник*. 2011. № 5. С. 179–183.
9. Hyndman R. J., Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*. 2008. Vol. 27. No. 3.
10. Лукашин Ю. П. Адаптивні методи короткострокового прогнозування часових рядів: учбовий посібник. Москва: Фінанси і статистика, 2003. 416 с.
11. Herrera F., Lozano M., Verdegay J. Tackling real-coded Genetic algorithms: operators and tools for the behaviour analysis. *Artificial Intelligence Review*. 1998. Vol. 12. No. 4. P. 265–319.
12. YAHOO! Finance. URL: https://finance.yahoo.com/ (дата звернення: 11.10.2019).
13. Kwiatkowski D., Phillips P.C., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*. 1992. Vol 54. P. 159–178.
14. Armstrong J., Adya M., Collopy F. Rule-Based Forecasting Using Judgment in Time Series Extrapolations. *Principles of Forecasting: A Handbook for Researcher and Practitioners*. 2001. Vol. 1. P. 259-282.
15. Wichard J. D, Ogorzalek M. Time series prediction with ensemble models. *Neurocomputing*. 2007. Vol. 2. P. 2371–2378.