MARITIME TECHNICAL JOURNAL

2022 1 (224)

sciendo

DOI:10.2478/sjpna-2022-0003

ARIMA MODEL OPTIMAL SELECTION FOR TIME SERIES FORECASTING

Vadim Romanuke

Polish Naval Academy, Faculty of Mechanical and Electrical Engineering, Śmidowicza 69 Str., 81-127 Gdynia, Poland; e-mail: v.romanuke@amw.gdynia.pl; ORCID ID: 0000-0003-3543-3087

ABSTRACT

A fast-and-flexible method of ARIMA model optimal selection is suggested for univariate time series forecasting. The method allows obtaining as-highly-accurate-as-possible forecasts automatically. It is based on effectively finding lags by the autocorrelation function of a detrended time series, where the best-fitting polynomial trend is subtracted from the time series. The forecasting quality criteria are the root-mean-square error (RMSE) and the maximum absolute error (MaxAE) allowing to register information about the average inaccuracy and worst outlier. Thus, the ARIMA model optimal selection is performed by simultaneously minimizing RMSE and MaxAE, whereupon the minimum defines the best model. Otherwise, if the minimum does not exist, a combination of minimal-RMSE and minimal-MaxAE ARIMA models is used.

Key words:

time series forecasting, ARIMA, model selection, forecasting horizon, seasonality, trend.

Research article

INTRODUCTION

Time series analysis and forecasting is an important field which has both deep theoretical and practical impact. In this field, a sequence of practically influential values registered at discrete time steps is referred to as a time series and is used to forecast future values. The forecasting accuracy depends on a model, which is used to generate forecasts, and the forecasting horizon. As the horizon is extended, the accuracy decreases. Selection of a model to forecast a time series expansion is crucial to achieve the best accuracy, whichever the forecasting horizon is [13, 3, 8].

In analyzing a time series, the main difficulties are possible seasonality and trend effects. These effects are handled by autoregressive integrated moving average (ARIMA) approach [3, 8]. To forecast a time series, an ARIMA model is built by specifying three parameters: a degree of the nonseasonal autoregressive polynomial (NSARP), a degree of the nonseasonal differencing lag operator polynomial (NSDLOP), and a degree of the nonseasonal moving average polynomial (NSMAP). The NSARP degree can be defined by using properties of the autocorrelation function (ACF). A positive integer value of the NSDLOP degree corresponds to a trend in the time series. For instance, is the NSDLOP degree is set at 1, then it is assumed that the trend is linear. The NSMAP degree is adjusted in a trickier way. It requires iterative nonlinear fitting procedures by additional assumptions that random shocks are normally distributed by zero mean and constant scale [6].

A methodology from Box and Jenkins [2, 9] exists for ARIMA model optimal selection. Basically, the methodology consists in the model general structure identification, estimation of its parameters, and checking whether the estimated model conforms to the specifications of a stationary univariate process [2]. The last step of the methodology, at which stationarity is checked, is argued to be fundamentally problematic. The reason behind this lies in that real-world practically usable time series are never stationary. In the desired automation of time series forecasting, therefore, the question of ARIMA model optimal selection is still open to arguing.

MOTIVATION AND GOAL

Reasoning from that, in the general case, selecting an ARIMA model is weakly defined or undefined, the goal is to suggest a fast-and-flexible method of

ARIMA model optimal selection for univariate time series forecasting. The method should be independent of additional assumptions allowing to obtain best forecasts automatically. To achieve the goal, the selection criteria are to be substantiated, whereupon the ARIMA forecasts are automatically generated. The ARIMA model optimal selection will be tested on a set of benchmark time series. Finally, the results will be discussed and brief conclusions with an outlook for further research will be made.

ARIMA MODEL OPTIMAL SELECTION

Denote a degree of the NSARP by p, a degree of the NSDLOP by d, and a degree of the NSMAP by q. As it is classically denoted [2, 1], an ARIMA model is defined by triplet $\{p, d, q\}$. However, a lot of experimental data allow affirming that ARIMA models by $\{p, d, 0\}$ are not worse than ARIMA models by $\{0, d, q\}$. Moreover, ARIMA models by $\{p, d, 0\}$ are determined faster. Meanwhile, experiments show that setting both p and q at nonzero integers generate less accurate forecasts. Therefore, it is practically efficient to consider only ARIMA models defined by $\{p, d, 0\}$.

Denote by T_0 the amount of a time series data, which are used to select an ARIMA model. The data are formally denoted by

$$\{y(t_i)\}_{i=1}^{T_0}, \tag{1}$$

where, without losing generality, $t_i = i$. Data (1) can be also referred to as the time series. Let T_{forec} be a forecasting horizon, which is a time point to which forecasts are made, where $T_0 < T_{\text{forec}}$. Data

$$\{y(t_i)\}_{i=T_0+1}^{T_{\text{forec}}} \tag{2}$$

will be used for testing the forecasting quality (performance). Data (2) can be also referred to as the test data.

First, the time series (1) is approximated by a polynomial trend model

$$\bar{y}_{\text{trend}}(t) = \sum_{j=0}^{N} \beta_j t^j, \tag{3}$$

where $N = 1, 2, 3, \ldots$, successively. If the root-mean-square error (RMSE) [13, 8] of the polynomial trend model of degree $N+1$ is greater than the RMSE of the polynomial trend model of degree N, then the latter is presumed to be the time series trend. Then the ACF of the sequence

$$\left\{ y(t_i) - \bar{y}_{\text{trend}}(t_i) \right\}_{i=1}^{T_0} \tag{4}$$

is found (e. g., see [12, 5, 10]). All different lags in sequence (4) are determined by this ACF. For each lag, denoted by p_z, an ARIMA model is defined and determined by $\{p_z, N, 0\}$. The performance is estimated by the corresponding RMSE (not to be confused with the RMSE of the approximating trend) and the maximum absolute error (MaxAE) [4, 7, 11] as follows. If

$$\left\{ \tilde{y}(t_i) \right\}_{i=T_0+1}^{T_{\text{forec}}} \tag{5}$$

are forecasted data, they are normalized with respect to the initial data:

$$\tilde{u}(t_i) = \frac{\tilde{y}(t_i) - \min\limits_{k=\overline{T_0+1, T_{\text{forec}}}} y(t_k)}{\max\limits_{k=\overline{T_0+1, T_{\text{forec}}}} y(t_k) - \min\limits_{k=\overline{T_0+1, T_{\text{forec}}}} y(t_k)} \quad \text{by} \quad i = \overline{T_0 + 1, T_{\text{forec}}}. \tag{6}$$

Test data (2) are normalized as well:

$$u(t_i) = \frac{y(t_i) - \min\limits_{k=\overline{T_0+1, T_{\text{forec}}}} y(t_k)}{\max\limits_{k=\overline{T_0+1, T_{\text{forec}}}} y(t_k) - \min\limits_{k=\overline{T_0+1, T_{\text{forec}}}} y(t_k)} \quad \text{by} \quad i = \overline{T_0 + 1, T_{\text{forec}}}. \tag{7}$$

Then the RMSE is calculated as

$$\rho_{\text{RMSE}}(p_z, N) = \sqrt{\frac{1}{T_{\text{forec}} - T_0} \sum_{i=T_0+1}^{T_{\text{forec}}} \left[u(t_i) - \tilde{u}(t_i) \right]^2} \tag{8}$$

and the MaxAE is calculated as

$$\rho_{\text{MaxAE}}(p_z, N) = \max_{i=\overline{T_0+1, T_{\text{forec}}}} \left| u(t_i) - \tilde{u}(t_i) \right|. \tag{9}$$

MaxAE (9) registers information about the worst outlier. RMSE (8) and MaxAE (9) are used to see the averaged and worst errors in forecasting.

The best ARIMA model is selected by such a lag, at which RMSE (8) and MaxAE (9) are simultaneously minimal. If there are two different lags $p_z^{(1)}$ and $p_z^{(2)}$, at which

$$\rho_{\text{RMSE}}\left(p_z^{(1)}, N\right) < \rho_{\text{RMSE}}\left(p_z^{(2)}, N\right) \tag{10}$$

and

$$\rho_{\text{MaxAE}}\left(p_z^{(1)}, N\right) > \rho_{\text{MaxAE}}\left(p_z^{(2)}, N\right), \tag{11}$$

or

$$\rho_{\text{RMSE}}\left(p_z^{(1)}, N\right) > \rho_{\text{RMSE}}\left(p_z^{(2)}, N\right) \tag{12}$$

and

$$\rho_{\text{MaxAE}}\left(p_z^{(1)}, N\right) < \rho_{\text{MaxAE}}\left(p_z^{(2)}, N\right), \tag{13}$$

then the best ARIMA model is a combination of models determined by $\left\{p_z^{(1)}, N, 0\right\}$ and $\left\{p_z^{(2)}, N, 0\right\}$. The forecasts by this model are made as the average of the forecasts by models determined by $\left\{p_z^{(1)}, N, 0\right\}$ and $\left\{p_z^{(2)}, N, 0\right\}$.

RESULTS OF TESTING THE METHOD

To test the ARIMA model optimal selection method, a set of benchmark time series is formed. A benchmark time series is generated in the form of

$$y(t_i) = \left[a_1 + 0.25\Theta_1\left(T_{\text{forec}}\right)\right] r(t_i) + a_2\Theta_2\left(T_{\text{forec}}\right) + a_3 t_i + a_4 t_i^2 + a_5 t_i^3 \text{ by } i = \overline{1, T_{\text{forec}}} \tag{14}$$

where $a_1 = 2$, $\Theta_1\left(T_{\text{forec}}\right)$ and $\Theta_2\left(T_{\text{forec}}\right)$ are two vectors of T_{forec} pseudorandom numbers drawn from the standard normal distribution (with zero mean and unit variance), $\left\{r(t_i)\right\}_{i=1}^{T_{\text{forec}}}$ is a sequence of identical randomly-structured subsequences (IRSS) of periodicity $P_r \in \{6, 7, 8\}$, $a_2 = 0.175$, and

$$a_3 \in \left(-10^{-4}; 10^{-4}\right),\ a_3 \in \left(-10^{-5}; 10^{-5}\right),\ a_3 \in \left(-10^{-6}; 10^{-6}\right) \tag{15}$$

are randomly generated coefficients of the trend polynomial.

Obviously, the forecasting performance worsens as the forecasting horizon is extended. Having generated 200 benchmark time series by (14), (15), $T_0 = 84$ and $T_{\text{forec}} = 168$, a lot of extremely inaccurate forecasts are obtained (see fig. 1, where time series points are marked as dots, and forecasts are marked as squares).

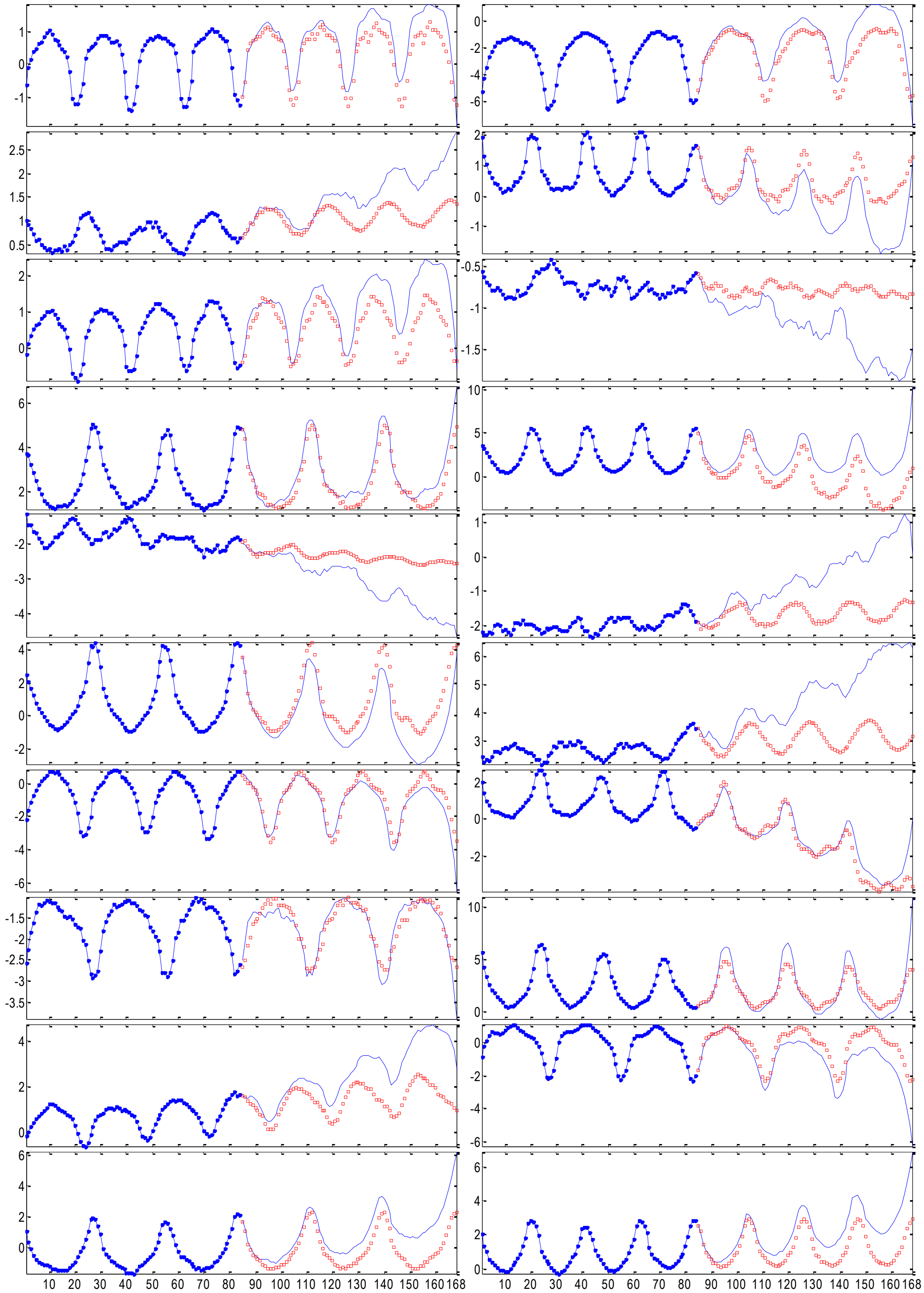

Fig. 1. A set of 20 time series forecasts whose horizon is the doubled length of the time series

It is clearly seen that the most inaccurate forecasts are typical for time series which have poorly visible periodicity. The distribution of RMSE (8) for the best ARIMA models of 200 time series by (14), (15), $T_0 = 84$, $T_{\text{forec}} = 168$ is presented in fig. 2. This distribution shows that there are many time series (among those remaining 180 instances) whose forecasts are as poor as those for time series ## 6, 9, 10, 12 in the respective subplots of fig. 1. Forecasts for time series #100 and #170 are the poorest. This is confirmed by fig. 3 presenting the distribution of MaxAE (9) for this case. Having compared the distributions of RMSE (8) and MaxAE (9), it is worth noting that they do not strongly correlate. Therefore, these criteria of ARIMA model selection can be thought of as complementary.

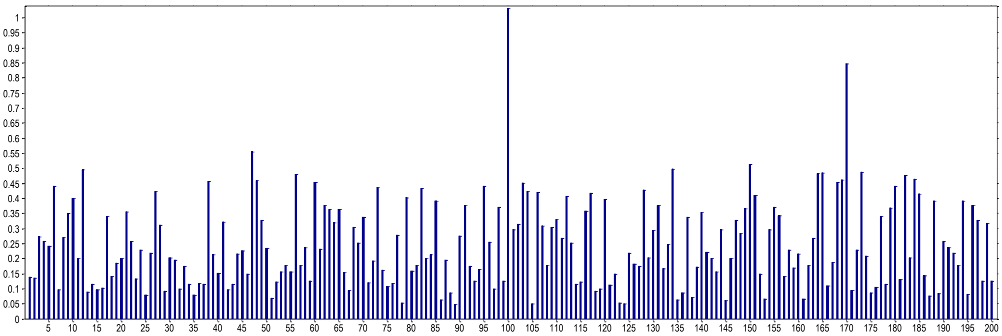

Fig. 2. RMSE (8) for the best ARIMA models of 200 time series by (14), (15), $T_0 = 84$ and $T_{\text{forec}} = 168$ (fig. 1)

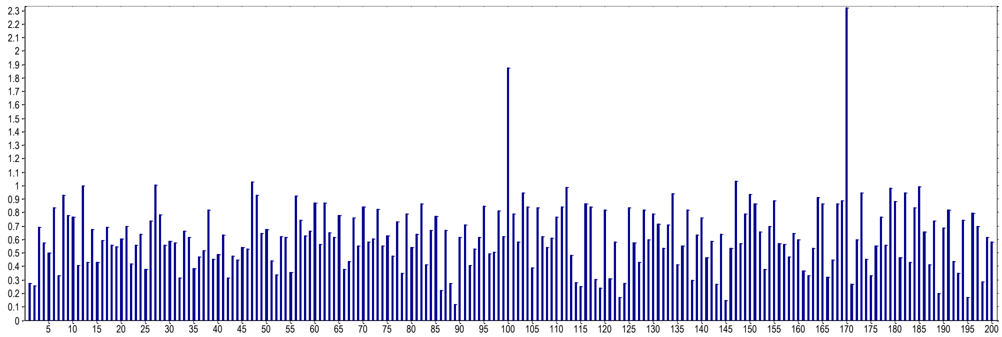

Fig. 3. MaxAE (9) for the best ARIMA models of 200 time series by (14), (15), $T_0 = 84$ and $T_{\text{forec}} = 168$ (fig. 1)

As the horizon length is shortened from 84 to 63 and 42, the ARIMA performance significantly improves (see fig. 4, where forecasts are marked as squares and stars, respectively). However, there are still cases, in which the forecasting

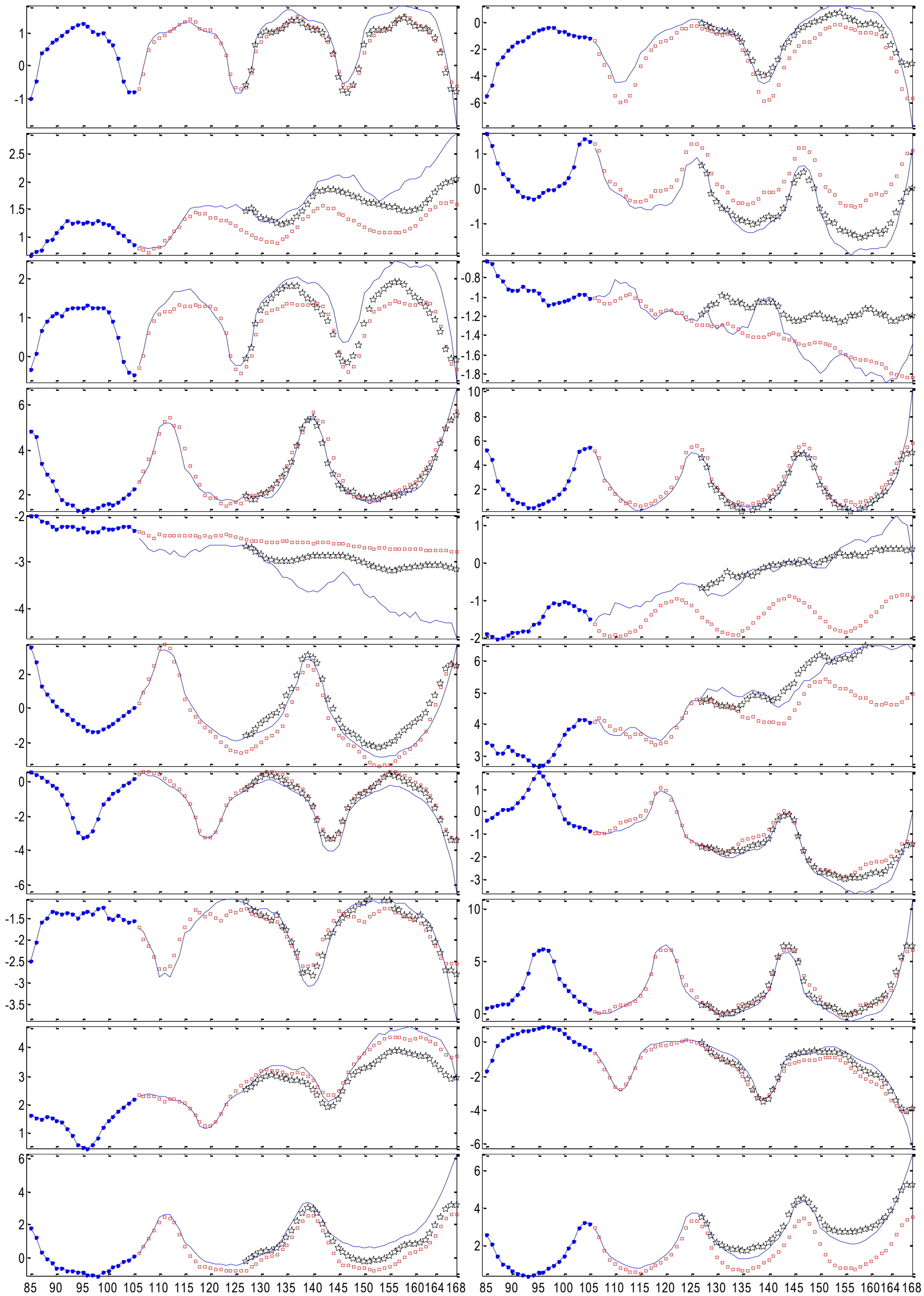

Fig. 4. The set of 20 time series (fig. 1) forecasts by shortening the horizon length to 63 and 42

accuracy for the shortest horizon length (i. e., by $T_0 = 126$) is worse than that for $T_0 = 105$. An example of this is seen in fig. 4 for time series #6 (the third upper subplot in the right column). Nevertheless, shortened-horizon forecasts mostly have become sufficiently accurate. For instance, forecasts for time series #100, unlike the case by $T_0 = 84$, have become a way better (fig. 5), as well as forecasts for time series #170 (fig. 6). The number of cases in which the forecasting accuracy has been improved after shortening the horizon length in presented in tab. 1. Surely, this is not a perfect result, but the result in tab. 2 is more convincing: the amount of

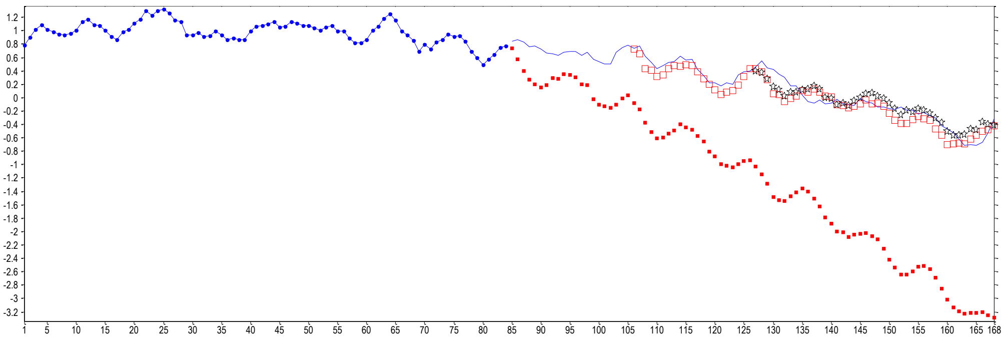

Fig. 5. Three cases of ARIMA forecasts for time series #100

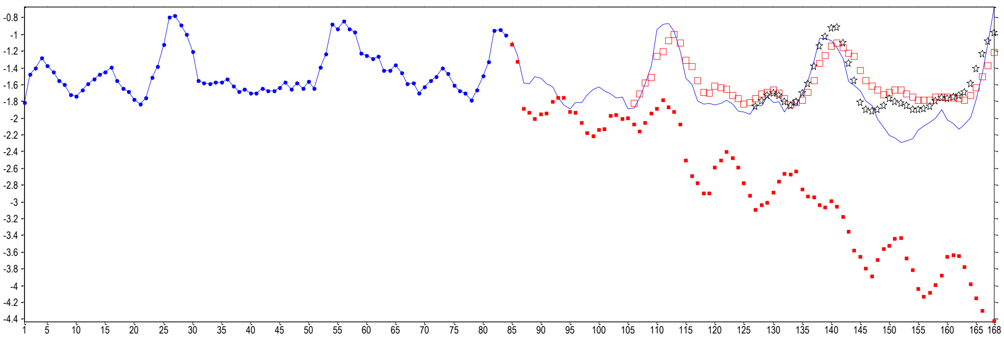

Fig. 6. Three cases of ARIMA forecasts for time series #170

Tab. 1. The number of cases (bold) in which shortening the horizon length improves the accuracy

RMSE				MaxAE			
preceding and succeeding length of the time series (T_0)							
84, 105	105, 126	84, 126	84, 105, 126	84, 105	105, 126	84, 126	84, 105, 126
149	**113**	**166**	**72**	**153**	**127**	**174**	**89**

Tab. 2. The number of cases (bold) in which RMSE and MaxAE do not exceed a threshold

	RMSE is less than 0.25			MaxAE is less than 0.5		
length of the time series (T_0)	84	105	126	84	105	126
The number of cases (out of 200)	**115**	**160**	**172**	**61**	**122**	**145**

sufficiently accurate forecasts has increased since 86 % of RMSE (see fig. 7) and 72.5 % of MaxAE (see fig. 8) do not exceed the acceptable threshold.

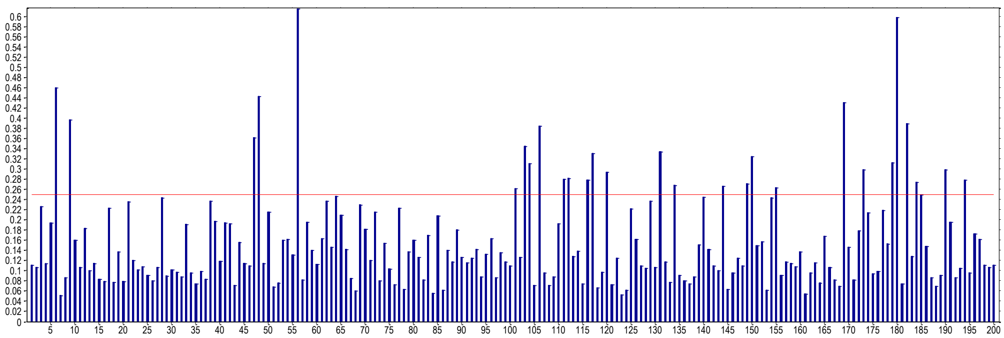

Fig. 7. RMSE (8) for the best ARIMA models of 200 time series by (14), (15), $T_0 = 126$ and $T_{\text{forec}} = 168$ (see star-marked forecasts in fig. 4)

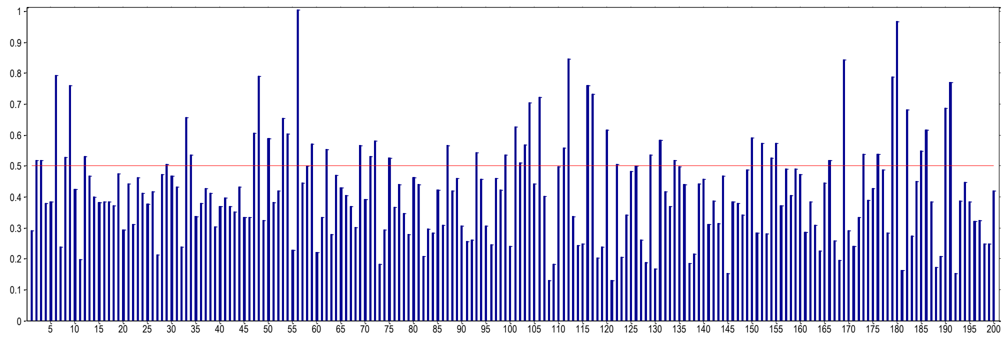

Fig. 8. MaxAE (9) for the best ARIMA models of 200 time series by (14), (15), $T_0 = 126$ and $T_{\text{forec}} = 168$ (see star-marked forecasts in fig. 4)

Among those 200 best ARIMA models obtained for the worst case (see fig. 1), where the horizon length is equal to the length of the time series, only 18 time series have been forecasted by combinations of two ARIMA models. This number has increased up to 45 and 55 for the cases with $T_0 = 105$ and $T_0 = 126$ (i. e., for the two longer time series), respectively. It is noteworthy that, in the worst benchmark case, 172 ARIMA models have been determined by a linear trend, i. e., by $N=1$. For the two longer time series with $T_0=105$ and $T_0=126$, the number of linear-trend ARIMA models has dropped to 106 and 99, respectively.

DISCUSSION

The suggested method does not select an appropriate forecasting horizon length. This can be considered as a drawback or concession. In real-world practical tasks, however, the horizon length is usually adjusted by studying short-length forecasts first.

Confidence intervals for ARIMA model forecasts stretch as the forecasting horizon length is made longer. Such intervals are based on assumptions that residuals are uncorrelated and normally distributed. If either of these assumptions does not hold, then the forecast intervals may be incorrect [3, 8]. This is why confidence intervals are not estimated for the 200 benchmark time series forecasts.

Obviously, the forecasting quality (which can be also referred to as the performance or accuracy) depends on the forecasting horizon length. The longer this length is, the poorer forecasts may become (see fig. 4 — 6), if even ARIMA models are selected optimally. Besides, the exemplary forecasts in fig. 1 allow claiming that time series with weakly visible periodicity are expected to be forecasted poorly.

The obtained results also allow claim that if the trend is linear or parabolic, ARIMA models perform forecasts more accurately. Time series with the cubic trend seem to be forecasted the worst. This is why the suggested method initially tries to forecast by assuming that the trend is linear, where $d=1$ is set. It is worth noting that even if the time series does not have a trend, the initial degree of the NSDLOP is set at $d=1$. Then, anyway, sequence (4) does not significantly differ from time series (1).

CONCLUSIONS

In order to improve the quality of forecasting, the suggested method automatically selects an ARIMA model or a combination of ARIMA models. The respective forecasts are close to be as accurate as possible owing to effectively finding lags and using diverse criteria of the forecasting quality. Lags are found by the ACF of a detrended time series, where the best-fitting polynomial trend is subtracted from the time series. The forecasting quality criteria are RMSE and MaxAE allowing to register information about the average inaccuracy and worst outlier. Thus, the ARIMA model optimal selection is performed by simultaneously minimizing RMSE and MaxAE, whereupon the minimum defines the best model. Otherwise, if the minimum does not exist, a combination of minimal-RMSE and minimal-MaxAE ARIMA models is used.

The research is to be furthered by considering a possibility to predict poor accuracy in forecasts before applying an approach. The poorly visible periodicity is an option (see fig. 1). Such a possibility, for instance, can optimize a forecasting horizon length. Obviously, this is a meta-forecasting methodology, and it will encompass much wider and deeper area of univariate time series.

REFERENCES

[1] Box G., Jenkins G., Reinsel G., *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ, 1994.

[2] Box G., Jenkins G., *Time Series Analysis: Forecasting and Control*, Holden-day, San Francisco, 1970.

[3] De Gooijer J. G., Hyndman R. J., *25 years of time series forecasting*, 'International Journal of Forecasting', 2006, Vol. 22, Iss. 3, pp. 443 — 473.

[4] Edwards R. E., *Functional Analysis. Theory and Applications*, Hold, Rinehart and Winston, 1965.

[5] Gubner J., *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge University Press, 2006.

[6] Hamilton J. D., *Time Series Analysis*, Princeton University Press, Princeton, NJ, 1994.

[7] Hyndman R., Koehler A., *Another look at measures of forecast accuracy*, 'International Journal of Forecasting', 2006, Vol. 22, Iss. 4, pp. 679 — 688.

[8] Kotu V., Deshpande B., *Data Science (Second Edition)*, Morgan Kaufmann, 2019.

[9] Pankratz A., *Forecasting with Univariate Box — Jenkins Models: Concepts and Cases*, John Wiley & Sons, 1983.

[10] Papoulis A., Probability, *Random variables and Stochastic processes*, McGraw-Hill, 1991.

[11] Romanuke V. V., *A minimax approach to mapping partial interval uncertainties into point estimates*, 'Journal of Mathematics and Applications', 2019, Vol. 42, pp. 147 — 185.

[12] Romanuke V. V., *Computational method of building orthogonal binary functions bases for multichannel communication systems with code channels division*, Mathematical Modeling and Computational Methods, Ternopil State Technical University, Ternopil, Ukraine, 2006.

[13] Schelter B., Winterhalder M., Timmer J., *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, Wiley, 2006.

OPTYMALNY DOBÓR MODELU ARIMA DLA PROGNOZOWANIA SZEREGÓW CZASOWYCH

STRESZCZENIE

W pracy zaproponowano szybką i elastyczną metodę optymalnego doboru modelu ARIMA na potrzeby prognozowania szeregów czasowych z jedną zmienną. Metoda pozwala na uzyskanie możliwie najdokładniejszych prognoz, opierając się na skutecznym znajdowaniu opóźnień. Poszukiwanie opóźnień realizowane jest za pomocą funkcji autokorelacji szeregu czasowego bez trendu, w którym najlepiej dopasowany trend wielomianowy jest odejmowany od szeregu czasowego. Za kryteria jakości prognozowania przyjęto średni błąd kwadratowy (RMSE) i maksymalny błąd bezwzględny (MaxAE), które pozwoliły na rejestrację informacji o średniej i maksymalnej niedokładności. Optymalny dobór modelu ARIMA odbywa się poprzez jednoczesną minimalizację RMSE i MaxAE, dla której wartość minimalna określa najlepszy model. W przeciwnym razie, jeśli minimum nie istnieje, używana jest kombinacja modeli ARIMA z minimalnym RMSE i minimalnym MaxAE.

Słowa kluczowe:

prognozowanie szeregów czasowych, ARIMA, dobór modelu, horyzont prognozowania, sezonowość, trend.