FORECASTING METHODS IN MODERN ENTERPRISE MANAGEMENT

Dariusz Grzesica[1]

Abstract

The aim of the chapter is to describe advanced methods of forecasting used in modern, high-tech enterprises. One of them – the ARMA model considers the strong dependence between the individual observations, used for prediction of time series, characterized by high dynamics of change. Here will be explained the process of selection of its parameters, and method design of the model. Second – harmonic analysis – uses, in turn, the cyclicality of the time series by which to construct the model describing the time series and what is the forecast for future periods. The first method is a group of parametric models, the second one is nonparametric, and both use the nature of a change of time series. The use of these methods will be shown by example.

Keywords*: advanced forecasting methods, ARMA model, harmonic analysis.*

1. Introduction

Forecasting is an important element in the functioning of a modern enterprise and one of the main elements in management. It allows predicting the state that will come in the near future and to guide action to achieve the greatest benefit, often financial. Forecasting methods were developed for a long time. Their selection was dependent on phenomena that were to be predicted. As the development of industry, economy, and forecasting methods changed, they were adjusted to the specific conditions of the phenomenon, and new ones were created, which successfully set future values, characterized by a small forecast error. In addition, the development of modern information technologies has accelerated the computational process, and made the methods that remained in the field of theory.

Forecasting usually takes the nature of a complex process. In order to make a correct forecast, it is necessary to conduct appropriate measures and predictive.

The concept of prognostic approach is understood as a series of actions aimed at determining the future value of the phenomenon under investigation.

1 Dariusz Grzesica, Ph.D., Eng., Adiunkt, Institute of Road and Railway Engineering, Department of Transportation Systems, Cracow University of Technology, Warszawska street 24, 31-155 Krakow, e-mail: dgrzesica@pk.edu.pl.

For this predictive process, you can select several stages, which will increase the accuracy of the indicated forecast, namely:

- preliminary data analysis;
- construction of the forecasting model;
- model evaluation;
- determination of the predictive algorithm;
- evaluation of forecast quality (Bielinska, 2007).

Preliminary data analysis allows the selection of an appropriate forecasting method. One of the first methods based on time series, which was used in the enterprise was the exponential smoothing. They are identified by historical patterns, trends, or size of the seasonal and extrapolation of these patterns to future predicted values (Hirschey, 2009). The well-known is the Brown model. In 1944, Brown received the task to create a model for the detection of enemy submarines (Mills, 2011). The assignment of weights characterizes this model to the time series data. When the data is separated in time, the assigned weight disappear exponentially (Grzesica, 2015). For time series smoothing, in which there are random fluctuations, the Holt model is used. On the other hand, in the time series with the tendency of development, seasonal and random fluctuations the Winters model is used. The last two models are considered together as a Holt-Winters model.

The data used for prediction in the enterprise are temporary and persisted through the time series. Thus, it is possible to assign a variable at a specific point in time in which it was made, and the number of such variables to examine for a dependency that occurs between them.

Modern requirements for methods of forecasting are very high. They are taking into account the nature of the variable changes in the series and trying to read the reasons for these changes. In the information systems of the company, the modules for forecasting often used methods of exponential smoothing that considers the specific features of the time series, i.e., trend, seasonality, and fluctuations. More advanced methods, such as autoregressive and moving average models, can accurately contribute to determining future volumes with high predictability.

2. ARMA – forecasting models based on time series

Data is recorded on the enterprise demand, sales volume, material requirements, purchases, or costs values, which are characterized by high dynamics of changes. In this regard, the use of exponential smoothing models does not give satisfactory results. ARMA models are well proven in similar time series as they reflect the dependencies that occur between variables in a different period. Their main characteristic is the fact that the value of the forecasted variable at time t is a linear combination of values of the same variable from the previous periods *t - 1,t -2, …, t - p* plus a certain value in a random component (Grzesica & Więcek, 2014). ARMA models consist of two groups:

1) AR – autoregressive models
2) MA – moving average models

The autoregressive AR model is as follows:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t \tag{1}$$

where:

y_t, y_t-1, y_{t-2}, y_{t-p} – the value of the predicted variable at time *t, t* - 1, *t* - 2,…, *t* - *p*,
φ_0, φ_1, …, φ_p – model parameters,
e_t – the rest of the model in period t,
p – the value of delay.

The model assumes that there is autocorrelation between the values of the predicted variable and its delay values in time (Dittmann, 2003). The problem of determining the autoregressive order *p* is important and not always easy to solve (Maciąg et al., 2013).

Regardless of the accepted model of auto regression for the process occurring at the enterprise are affected, to varying degrees, by a random factor. The method using the random factor is called the moving average model MA based on the assumption that the value of the variable in the predicted period will be equal to the arithmetic average of the actual data for the last several periods (Reszka, 2010). It is expressed as:

$$y = \Phi_0 - \Phi_1 e_{t-1} - \Phi_2 e_{t-2} - \cdots - \Phi_q e_{t-q} + e_t \tag{2}$$

where:

e_t, e_{t-1}, e_{t-2}, e_{t-p} - the rest of the model in period *t, t* - 1, *t* - 2, *t* - *q*,
Φ_0, Φ_1, …, Φ_q - model parameters,
q - the value of delay.

To achieve a greater fit of the model for a time series sometimes requires the connection of both (AR) p and (MA) q models in one autoregressive and moving average model called ARMA with parameters (*p, q*). It is expressed by a formula:

$$y = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \mathrm{e_t} - \Phi_0 - \Phi_1 e_{t-1} - \Phi_2 e_{t-2} - \cdots - \Phi_q e_{t-q} + e_t \tag{3}$$

Difficulties in building the ARMA model causes the selection of its parameters. For their correct definition, it is necessary to perform the following steps:

- identification;

- estimation;
- verification.

In the identification process of the time series, a necessary condition for the application of ARMA model is to determine its stationarity, i.e., constant in time mean, variance, and autocorrelation of the studied series. For stationary time series testing the unit root, test developed by Dickeya and Fuller in 1979, is used most often (Dickey & Fuller, 1979). At that moment, when the level of significance of the parameter for the studied time series is higher than 0.05, then we are dealing with the non-stationarity of the series. In order to eliminate that condition, a differentiation method is used, namely the calculation of further differences according to the formula: $\Delta yt= yt - yt\text{-}1$. It is used until the time series is stationary (in practice no more than three times). Alternatively, the Schwarz's (Schwarz, 1978), Akaike's (1974) or Hannan-Quinn (1979) information criteria can be used. In case of removing non-stationarity from the time series, the ARMA model is expanded to ARIMA (*p, d, q*) (Autoregressive Integrated Moving Average), Containing parameter d, which is equal to the number of time series differentiation processes. The parameters *p* and *q* are determined from autocorrelation and partial autocorrelation.

The process of model estimation is to determine its parameters (*p, d, q*). This can be done using the method of least squares, a method of maximum likelihood or using the equations of Yule – Walker (Box & Jenkins, 1983). The model parameters are determined based on ACF autocorrelation (*Autocorrelation Function*) and PACF partial autocorrelation (*Partial Autocorrelation Function*). This is done using the Box-Jenkins method (*Box & Jenkins, 2015*). The farther into the past, the smaller the influence of the time series variables exerts on the present.

Verification of an ARIMA model is performed through the study of the autocorrelation of the residues of the model, i.e., the difference between the values of the accepted model, and the actual. The model shows no autocorrelation of residuals in the case when the coefficients are not significantly different from zero. Otherwise, it is necessary to choose the model parameters again.

3. Harmonic analysis – forecasting based on cyclical changes in the time series

Harmonic analysis, called spectral analysis, is a method that uses the theory and application of Fourier series in respect of the studied phenomenon. Most of the studied phenomenon covers more than a year. This method is widely used in astronomy to study stars and in chemistry to study the chemical composition of substances. There are two conditions which occurring together or separately allow the application of spectral analysis:

- the studied period is more than one year;
- a sufficiently large amount of data enables the cyclicity of the time series to be identified.

The second condition allows the study of time series for the period less than one year. The question remains, how short the period may be possible to use spectral analysis. It seems that there is no limitation. The only limitation is the accuracy of the forecast achieved with the help of this analysis. It can be concluded that as a result of large amounts of data, cyclicity is always present, in greater or lesser degree. Therefore, the application of harmonic analysis for forecasting purposes in the enterprise is justified, given the fact that a significant part of the processes taking place inside and in the closed environment of the enterprise has a periodic character. This periodicity may be associated with the seasons (seasonality) or certain days of the week (the syndrome of the day – for example, the beginning and the end of the week). Sometimes the cyclical nature appears at intervals difficult to explain for the prognostician. This does not mean, however, that it does not occur.

Harmonic analysis of time series is a kind of analysis in which a discrete Fourier transform (DFT) transforms a time series into a frequency domain. It consists of building a model in the form of the sum of the so-called harmonics, i.e., sinusoidal and/or cosineal functions of a given period (Zelias et al., 2003). The first harmonic has a period equal to the duration of the study period, the second – a half of this period, the third – one-third of it. In the case of n observations, the number of all possible harmonics is n/2.

A condition for the application of harmonic analysis is the determination of the spectral density function (spectrum), which allows to identify the harmonic structure of the time series and to determine the impact of the individual components of the formation process variance (Talaga & Zieliński, 1986). Spectrum is represented in a limiting way, where the abscissa axis represents the periods or frequencies and along the ordinate axis an estimate of the spectral density function.

To build the model based on the spectral analysis the following formula is used:

$$y_t = \alpha_0 + \sum^{n/2} \left(\alpha_{\mathrm{i}} \sin\left(\frac{2\pi}{n} \cdot i \cdot t\right) + \beta_i \cos\left(\frac{2\pi}{n} \cdot i \cdot t\right)\right) \tag{4}$$

where:
i – number of harmionics,
α_0, α_i, β_i – model parameters.

The model is applied at the moment of occurrence of fluctuations around a constant average level, represented by parameter α_0. In the case of trend, a formula is applied:

$$y_t = f(t) + \sum_{i=1}^{n/2} \left(\alpha_i \sin\left(\frac{2\pi}{n} \cdot i \cdot t\right) + \beta_i \cos\left(\frac{2\pi}{n} \cdot i \cdot t\right)\right) \tag{5}$$

where:
f(t) – trend function.

The model parameters are determined based on the least squares method in accordance with the formulas:

$$a_0 = \frac{1}{n}\sum_{t=1}^{n} y_t \tag{6}$$

$$a_i = \frac{2}{n}\sum_{t=1}^{n} y_t \cdot \sin\left(\frac{2\pi}{n} \cdot i \cdot t\right), \text{for } i = 1, \dots, \frac{n}{2} - 1, \tag{7}$$

$$b_i = \frac{2}{n}\sum_{t=1}^{n} y_t \cdot \cos\left(\frac{2\pi}{n} \cdot i \cdot t\right), \text{for } i = 1, \dots, \frac{n}{2} - 1, \tag{8}$$

where:

a_0, a_i, b_i – estimates of the parameters α_0, α_i, β_i.

The longer the time series, the greater may be the number of harmonics. For latest harmonic with number n/2, parameter an/2 is always zero, and the parameter bn/2 is determined by the formula:

$$b_{n/2} = \frac{1}{n}\sum_{i=1}^{n}[y_t \cos(\pi t)] \tag{9}$$

To determine the j-th harmonic and the contribution that a given harmonic influences on the variance of the predicted variable, the following formulas are most commonly used:

$$\omega_i = \frac{a_i^2 + b_i^2}{2\sigma^2} \quad dla\ i = 1 \dots \frac{n}{2} - 1 \tag{10}$$

$$\omega_i = \frac{a_i^2 + b_i^2}{\sigma^2} \quad dla\ i = \frac{n}{2} \tag{11}$$

where:
σ^2- variance of forecasted variable after a preliminary exception trends.

4. Verification of ARMA models and spectral analysis on the example of the number of orders in an enterprise

4.1. Forecasting based on ARMA model

Model and forecast based on real data coming from the forwarder enterprises, which is the number of orders realized in the process of cargo flow will be presented. Data taken from 2016, recorded 50 observations that were included in weekly periods is shown in Figure 1.

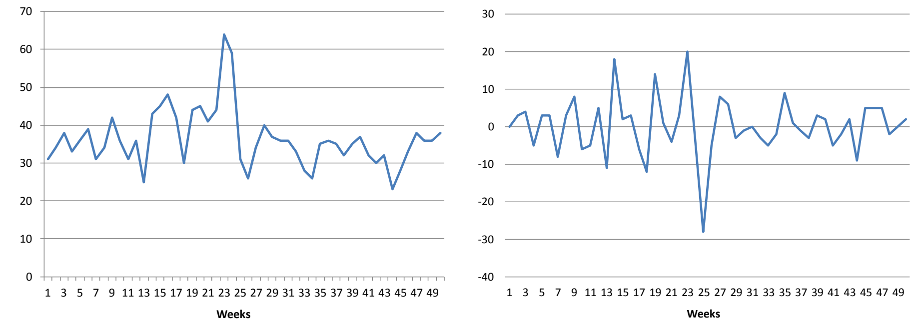

Figure 1. Number of orders in the forwarding company, a) non-stationary, b) stationary

In this example, 48 observations will be used to build the model, while the last two will be used to verify the prediction.

The first step in building the ARMA model is to check the stationarity of the time series. The calculated value of statistics of an Augmented Dickey-Fuller test (ADF). For the non-stationary series, $p = 0.6318$ and is greater than the critical value for the assumed significance level of $p = 0.05$. Thus, a time series of differentiation was made to eliminate the trend. The value of the ADF statistics is 0.00000967, and therefore the series is stationary.

In Fig. 2 shows the autocorrelation and partial autocorrelation of the studied time series. As in the case of autocorrelation and partial autocorrelation shows that the first collection of delay significantly affects the explanation of the variability of the time series; therefore, the parameters of the ARMA model was accepted at the level $p = 1$ and $q = 1$. The forecast is thus determined based on the model ARIMA (1, 1, 1), following the formula (3).

The model is presented as follows:

$$\mathrm{Y_t} = 0{,}4948Y_{t-1} - 1e_{t-1} \tag{12}$$

Validation of the ARMA model is through the study of the autocorrelation of the model residues. The model shows no autocorrelation of residuals in the case when the coefficients do not differ significantly from zero (Figure 2).

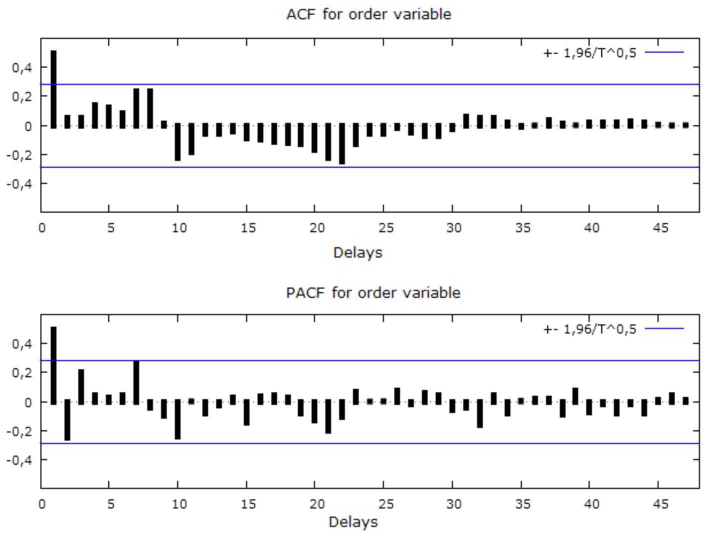

Figure 2. Autocorrelation (ACF) and partial autocorrelation (PACF) function

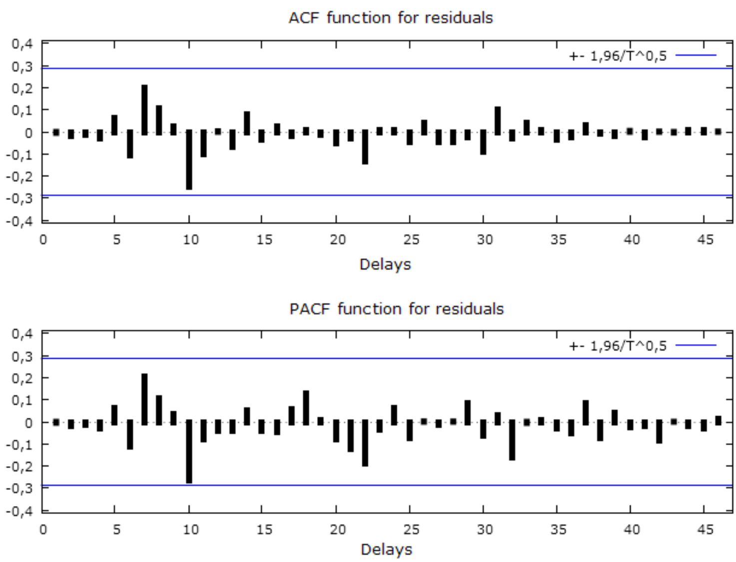

Figure 3. Autocorrelation (ACF) and partial autocorrelation (PACF) of residuals

In Figure 3 the graph shows no autocorrelation of the residues. In this regard, it is possible to construct forecast for future periods. Using the ARIMA (1, 1, 1) model, the forecast is as shown in Fig. 4. Of the 50 observations, 48 were taken to build the model, while others will serve as verification of the forecast.

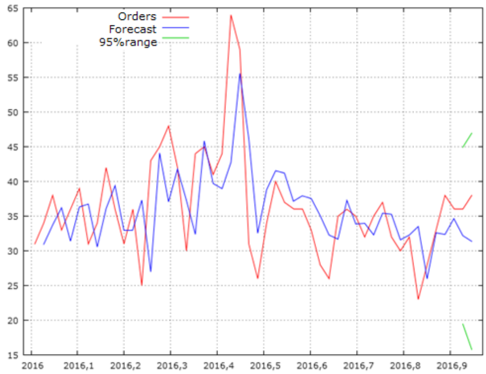

Figure 4. The real data with the model and forecast

Empirical data show great variability. It can be noted that the model well reproduces the real data and the forecast for the next period are close to real data.

4.2. Forecasting based on spectral analysis

The basis of harmonic analysis is to examine if the time series contains cyclical changes. For this purpose, a periodogram which is designed to detect cyclical fluctuations is used. Periodogram showing the number of orders is shown in Figure 5.

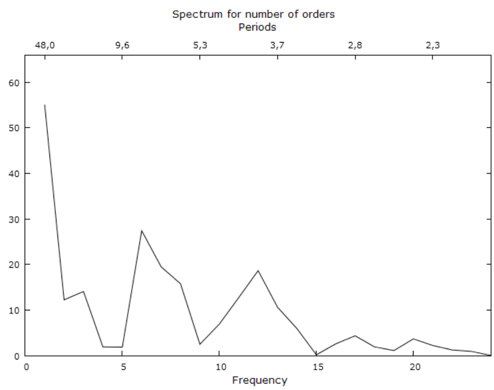

Figure 5. Periodogram of number of orders

On the basis of the periodogram, it can be concluded that the time series has cyclic fluctuations. Accordingly, harmonics will be designated in accordance with (10) and (11). The harmonics values that have the greatest share in the explanation of the series are shown in Table 1. The combined share of individual harmonics in explaining the volatility of the time series is almost 72%.

Table 1. Harmonics used to build the model

Number of harmonic	Designation	Period [week]	Share in the explanation [%]
6	n/6	8.00	6.38
7	n/7	6.86	8.01
8	n/8	6.00	6.22
11	n/11	4.36	11.34
12	n/12	4.00	15.46
13	n/13	3.69	12.17
14	n/14	3.43	6.34
17	n/17	2.82	6.02

Using the harmonics from Table 1 and the model parameters in the process of transforming the time series into the frequency domain, the model was constructed in the form:

$$y_t = 36{,}25 - 2{,}484 \cdot sin\left(\frac{2\pi}{48}t\right) - 1{,}073 \cdot cos\left(\frac{2\pi}{48}t\right) + 2{,}2 \cdot sin\left(\frac{2\pi}{48}t\right) + 2{,}087 \cdot cos\left(\frac{2\pi}{48}t\right) - 1{,}84 \cdot sin\left(\frac{2\pi}{48}t\right) - 1{,}938 \cdot cos\left(\frac{2\pi}{48}t\right) + 3{,}596 \cdot sin\left(\frac{2\pi}{48}t\right) - 0{,}3 \cdot cos\left(\frac{2\pi}{48}t\right) - 4{,}208 \cdot sin\left(\frac{2\pi}{48}t\right) - 0{,}167 \cdot cos\left(\frac{2\pi}{48}t\right) + 3{,}248 \cdot sin\left(\frac{2\pi}{48}t\right) - 1{,}848 \cdot cos\left(\frac{2\pi}{48}t\right) - 1{,}729 \cdot sin\left(\frac{2\pi}{48}t\right) - 2{,}071 \cdot cos\left(\frac{2\pi}{48}t\right) + 0{,}784 \cdot sin\left(\frac{2\pi}{48}t\right) + 2{,}508 \cdot cos\left(\frac{2\pi}{48}t\right) \quad (13)$$

On the basis of formula (13) the model, along with a forecast for two future periods (Fig. 6) was created.

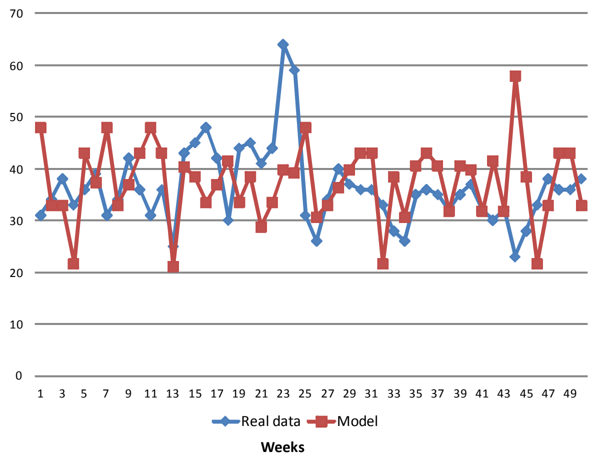

Figure 6. The real data with the model and forecast

5. Evaluation of forecast accuracy

The accuracy of the forecast on the basis of relative and absolute forecast errors (Table. 2) use ARMA models and spectral analysis will be considered.

Table 2. Indicators of forecast accuracy

Method	Forecast period	Real data	Forecast	Mean Absolute Error (MAE)	Mean absolute percentage terror (MAPE)
Spectral analysis	49	36	43.079	7.079	19.663
	50	38	32.907	5.093	13.402
ARIMA model (1,1,1)	49	36	32.19	3.81	10.583
	50	38	31.37	6.63	17.447
ARIMA model (1,1,2)	49	36	33.29	2.71	7.528
	50	38	33.44	4.56	12

Forecasts based on ARMA models and spectral analysis give forecasts at the accuracy level of 7.5 – 20%. For ARMA models the accuracy of the forecast depends on the skill and experience of the forecaster. The forecast for the first period for the ARMA (1, 1, 1) model is more than 10%, while for ARMA (1, 1, 2) model – 7.528 %. For the same period, the forecast based on spectral analysis is 19.67%. The greater predictive error for spectral analysis is probably due to insufficient lengths of time series (low number of observations) or the period in which the data are recorded (weekly periods).

The model would therefore not capture cyclical changes, which might be visible in shorter or longer periods of time.

6. Conclusions

Forecasting methods based on time series in the company should include the evaluation and extraction of the deterministic part of the series (trend, seasonality, randomness, cyclicity). In this regard, methods of modeling and forecasting time series must be properly chosen.

ARMA models are quite good at modeling variables that are characterized by high dynamics and strong correlation between observations. These observations provide some information about the evolution of successive series changes and thus allow for the prediction of high accuracy. The researcher's experience, which is selected based on many forecasts, plays a major role in the selection of parameters.

Models based on harmonic analysis, in turn, are excellent to use in time series characterized by cyclical changes. The spectrum shows that the series has cyclical fluctuations, which gives the possibility of modeling and forecasting future values based on harmonics that significantly influence the explanation of changes in time series.

Both methods are advanced, the computational process is very complex, and the estimation of some parameters is based on the experience of the forecaster. However, application of both methods seems to be necessary regarding the growing needs of modern enterprises in the field of forecasting.

References

Akaike, H. (1974). A New Look at the Statistical Model Identification. *I.E.E.E. Transactions on Automatic Control.* No.19.

Bielińska, E. (2007). *Prognozowanie Ciągów Czasowych.* Gliwice: Wydawnictwo Politechniki Śląskiej.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control.* New Jersey: John Wiley & Sons.

Box, G.E.P., & Jenkins, G.M. (1983). *Analiza Szeregów Czasowych. Prognozowanie i Sterowanie*. PWN, Warszawa.

Dickey, D.A., & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 427-431.

Dittmann, P. (2003). *Prognozowanie w przedsiębiorstwie. Metody i ich Zastosowanie.* Kraków: Oficyna Ekonomiczna.

Grzesica, D., & Więcek, P. (2014). Wykorzystanie modeli autoregresji i średniej ruchomej w prognozowaniu wielkości popytu niezależnego. *Czasopismo Logistyka 4*, 3908-3913.

Grzesica, D. (2015). Elementy zarządzania popytem w łańcuchu dostaw. In O. Gugnin, H. Hall & L. Witek (Eds.), *Zarządzanie Marketingowe Współczesnymi Organizacjami - Determinanty, Funkcje, Strategie.* Rzeszów: Wydawnictwo Politechniki Rzeszowskiej.

Hannan, E.J., & Quinn, B.G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, 41*(2), 190-195.

Hirschey, M. (2009). *Managerial economics, 12th editio.* London: Cangage Learning.

Krawczyk, S. (2001). *Metody Ilościowe w Planowaniu Działalności Przedsiębiorstwa*. Warszawa: Wydawnictwo C.H. Beck.

Maciąg, A., Pietroń, R., & Kukla, S. (2013). *Prognozowanie i Symulacja w Przedsiębiorstwie*. Warszawa: PWE.

Mills, T. (2011). *The Foundations of Modern Time Series Analysis*. Palgrave MacMillan,.

Reszka, L. (2010). *Prognozowanie Popytu w Logistyce Małego Przedsiębiorstwa.* Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.

Talaga, L., & Zieliński, Z. (1986). *Analiza Spektralna w Modelowaniu Ekonometrycznym*. Warszawa: PWN.

Zeliaś, A., Pawełek, B., & Wanat, S. (2003). *Prognozowanie Ekonomiczne: Teoria, Przykłady, Zadania*. Warszawa: Wydawnictwo Naukowe PWN.

Biographical note

Dariusz Grzesica, Assistant Professor in the Institute of Construction and Transportation Engineering & Management, Krakow University of Technology. Ph.D. in Economics, the field of Transport, Spedition and Logistics (TSL). Scientific interests focus on the problems of modeling and forecasting time series, also optimization of flow goods and transportation networks.