The Problem of the SARIMA Model Selection for the Forecasting Purpose

Josef Arlt[1] | *University of Economics, Prague, Czech Republic*
Peter Trcka[2] | *University of Economics, Prague, Czech Republic*
Markéta Arltová[3] | *University of Economics, Prague, Czech Republic*

Abstract

The goal of the work is to assess the ability to identify the proper models for the time series generated by SARIMA processes with different parameter values and to analyze the accuracy of the forecasts based on the selected models. The work is based on the simulation study. To this end, a new automatic SARIMA modelling method is proposed. Other competing automatic SARIMA modelling procedures are applied as well and the results are compared. The important question to which the reference should be made is the relation of the magnitude of the SARIMA process parameters i. e. the size of the systematic part of the process and the ability to identify a proper model. Another issue addressed herein is the relationship between the quality of the identified model and the accuracy of forecasts achieved by its application. The simulation study leads to the results that can be generalized to most empirical analyses in various research areas.[4]

Keywords	***JEL code***
SARIMA, simulation, identification of model, forecasting	*C15, C22, C63*

INTRODUCTION

The principle and the application of the SARIMA models in the time series modelling has been well known for many years. Its practical applications can be found in many areas where empirical analyses are needed and it has become a basis standard tool of modern econometric analysis. The crucial phase of the practical application of the Box-Jenkins methodology is the identification and verification of the suitable model.

The goal of this article is to find the time series for which it is relatively easy to identify the proper model and the time series for which it is difficult. Another goal is to analyze the forecasting abilities of the SARIMA models for different kinds of time series. A convenient way to verify the aforementioned is the simulation study and the application of the automatic SARIMA procedures.

---

[1] Department of Statistics and Probability, nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic. Corresponding author: e-mail: josef.arlt@vse.cz

[2] Department of Statistics and Probability, nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic. E-mail: trcp00@vse.cz

[3] Department of Statistics and Probability, nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic. E-mail: marketa.arltova@vse.cz

[4] This article is based on contribution at the conference *ASMDA 2017* – 17th conference of the Applied Stochastic Models and Data Analysis International Society and the Demographics 2017 Workshop.

The article is divided into four sections (excluding the Introduction). In the first section the SARIMA models are briefly described. In the second section, the simulation study as well as the Auto.SARIMA and Auto.AIC procedures for automatic model selection are explained. The results of the simulation study are the subject of the third section. The fourth section contains the conclusion, along with the summary of the work.

1 SARIMA MODELING AND FORECASTING

The ARMA(p, q) proces (Auto-Regressive-Moving-Average proces of orderes p, q) is defined as $\phi(B)y_t = c + \theta(B)a_t$, where B ($B^j y_t = y_{t-j}$) is the backshift operator and $\phi(B)$ and $\theta(B)$ are the polynomials in the lag operators of the order p and q respectively, $\{a_t\}$ is the white noise process. It is stationary, if the roots of the autoregressive polynomial $\phi(B)$ lie outside of the unit circle and it is invertible if the roots of the moving average polynomial $\theta(B)$ lie outside of the unit circle.

The SARMA(p, q)(P, Q)s proces (Seasonal ARMA process of orders p, q, P, Q) can be written in the form $\phi(B)\Phi(B^s)y_t = c + \theta(B)\Theta(B^s)a_t$, where s is the number of seasons (usually 4 or 12) and $\Phi(B^s)$ and $\Theta(B^s)$ are seasonal polynomials in the lag of the order P and Q respectively. It is denoted as SARMA(p,q)(P,Q)s. If the roots of all polynomials lie outside of the unit circle, the proces is stationary and invertible.

The special form of the non-stationary proces is the so called integrated process („I“ in acronym). Such a proces is stationary after some degree of differencing. The SARIMA(p,d,q)(P,D,Q)s proces is the general form of the integrated proces and can be written as $\phi(B)\Phi(B^s)\Delta^d\Delta_s^D y_t = c + \theta(B)\Theta(B^s)a_t$, where $\Delta^d = (1 - B)^d$ is the nonseasonal difference of the order d and $\Delta_s^D = (1 - B^s)^D$ is the seasonal difference of the order D.

The forecasting of the future values of the time series is an important role of the SARIMA modelling. The optimal forecast, i. e. the forecast with the minimum mean square error, is the conditional mean of the future random variable, which is conditioned on the historical information available in the observed values of the applied time series.

The SARIMA time series modelling methodology has been well known for many years and there exists a vast amount literature devoted to this topic, *inter alia*, Box, Jenkins, Reinsel and Ljung (2015), Brockwell and Davis (2010), Wei (2005), Hamilton (1994), Enders (2014), Pesaran (2016).

2 SIMULATION STUDY

The goal of the simulation study is to analyze the relationship of the magnitude of the SARIMA process parameters; i. e. the size of the systematic part of the process, which is used for time series generation and the ability to select the proper model for the generated time series. This question is general in scope, and the qualified and substantiated answers can be important for the empirical analyses in the different fields of the research. Another goal is to analyze the quality of the forecasts for the time series generated by the processes with different systematic parts. Important is also the analysis of the ability to select suitable model and reach the relatively accurate forecasts for the time series generated by the near non-stationary and the non stationary processes.

In the simulation study the results of the two automatic procedures for SARIMA model selection and forecasting are presented. The first one is based on the classic model selection process, i.e. the model identification, the parameters estimation, the diagnostic controll (on the basis of the residual time series, the autocorrelation, the heteroscedasticity as well as the normality are tested). The second one is based on the minimization of the AIC criterion (Akaike, 1974). Both procedures were implemented in the R software (2008).

2.1 Procedure Auto.SARIMA

The Auto.SARIMA is fully automated procedure, whose goal is to find the best model with respect to predefined parameters for the analyzed time series. In the first stage, the order of the nonseasonal and the seasonal differencing, i. e. the numbers d and D, after which the analyzed time series is stationary, has been found. For the nonseasonal unit root testing, the ADF (Dickey and Fuller, 1979), the PP (Phillips and Perron, 1988) and the KPSS (Kwiatovski, Phillips, Schmidt and Shin, 1992) tests are used. The seasonal unit root is tested by the CH test (Canova and Hansen, 1995).

The procedure will analyze the quality of the SARIMA(p,d,q)(P,D,Q)s models for the given order of the nonseasonal differencing d, as well as the seasonal differencing D, and for all possible combinations of values p, q, P, Q. It is therefore possible to skip the identification stage and to estimate the parameters for all the possible model forms. After the parameters estimation, the procedure continues with the diagnostic checking, which is mainly based on the residual analysis. The statistical significance of the parameter estimates is verified by the standard t tests. The autocorrelation is assessed by the residual autocorrelation function, and the Ljung-Box test (Ljung and Box, 1978). The conditional heteroscedasticity is tested by the ARCH test (Engle, 1982). The normality is tested by the Jarque-Bera test (Jarque and Bera, 1980).

If the parameter estimates are statistically significant and the null hypotheses of no autocorrelation, no conditional heteroscedasticity and normality are not rejected, then the value 1 is assigned to the particular property (autocorrelation, heteroscedasticity, normality, parameter significance). Otherwise, the value 0 is assigned. The suitability criterion of the model is computed as the weighted average of the results of the individual tests, where the individual properties have specific weight. The final value of each model is computed as a function of the value of the model suitability criterion and the value of the AIC criterion. The system mentioned above has been proposed by Trcka (2015).

2.2 Procedure Auto.AIC

The model selection on the basis of the AIC criterion is the content of the Auto.AIC procedure. The course of the procedure can be divided into four steps. In the first step, the stationarity of the time series is analyzed. The order of differencing is determined by the same methods as in the Auto.SARIMA procedure (see part 2.1). According to the order of differencing and the SARIMA model maximal orders, the set of the possible models is generated. Furthermore, the optimization criterion is set to such value which the AIC criterion cannot reach. In the third step, the adjustments are made so that all the models lead to the same number of residuals. On the basis of the adjusted time series, the model parameters are estimated, and the value of the AIC criterion is computed. In the following step, the actual value of the AIC criterion is compared with the value of the optimalization criterion. If the model is better that the last one, i. e. if its value of the AIC criterion is smaller than the value of the optimalization criterion, then it is denoted as the optimal model and the value of the optimization criterion is updated. In this manner the whole set of possible models is checked.

2.3 Data generation

In the simulation study, the time series generated by the SARIMA proces of the first order are analyzed. This process has the following form:

$$(1-\phi_1 B)(1-\Phi_1 B^{12})y_t = (1-\theta_1 B)(1-\Theta_1 B^{12})a_t. \tag{1}$$

The basic elements for the simulations are the time series generated by the normal white noise process with the variance $\sigma_a^2 = 1$. The parameters ϕ_1, θ_1, Φ_1, Θ_1 take all possible combinations of the following values: 0.0, 0.1, 0.2, 0.3, 0.4, 0,5, 0.6, 0.7, 0.8, 0.9, 1.0 (only positive parameters are used because in the economic practice, the SARIMA models with negative parameters occur rarely). When $\phi_1 = 1$, the process is non-seasonally non-stationary, when $\Phi_1 = 1$, the process is seasonally non-stationary, when $\phi_1 = 1$ and $\Phi_1 = 1$, the process is both non-seasonally and seasonally non-stationary. When $\theta_1 = 1$, the process is non-seasonally noninvertible, when $\Theta_1 = 1$, it is seasonally noninvertible or both, when $\theta_1 = 1$ and $\Theta_1 = 1$. Overall, the time series from 14 641 different generating processes are analyzed. Each process generates 100 time series with a length of 150 values. The time series generator was created in the R software.

3 RESULTS

The results of the simulation study are presented in a two-dimensional space, whose structure is shown in Table 1. The possible values for p, d, q, P, D, Q of the selected models are 0 or 1. The rows of table represent an ordered combination of values of the seasonal parameters Φ_1 and Θ_1 and the columns of table represent an ordered combination of values of the nonseasonal parameters ϕ_1 and θ_1. In this way the whole set of the all possible generating processes is arranged.

The table is conditionally formatted to be able to visually evaluate the results and success of the individual automatic procedures when comparing their ability to find a suitable model. This feature is referred to as quality criterion. The quality criterion can take the values in the interval from 0 to 100 and it represents the percentage success rate of the selection of the correct model by the given procedure.

The forecasts are computed as the point estimates of the conditional expectations of the future random variables. The analyzed time series with a length of 150 values, which is about 24 observations longer than the series used for model selection, is the input of this function. In the first step, the forecasts with the horizon $h = 24$ values are computed on the basis of the model estimated from the first 126 values. In the second step, the RMSE criterion is computed. The resulting RMSE value is computed as the average from the all partial RMSE values of 100 time series forecasts with a horizon of 24 values. This criterion is presented in the same way as the quality criterion.

Table 1 The Detail of Arrangement of Values in Table

	A	B	C	D	E	F	G
1							
2		XX	AR	0	0	0	0
3		SAR	SMA/MA	0	0,1	0,2	0,3
4		0	0	0,985798631	0,978848707	1,011897837	1,057034876
5		0	0,1	1,002931047	1,030341625	1,005169594	1,055862883
6		0	0,2	1,009372642	1,017211208	1,039043504	1,046414115
7		0	0,3	1,062359565	1,049209862	1,076806091	1,090812811
8		0	0,4	1,039299045	1,071779671	1,096205459	1,084753008
9		0	0,5	1,045760435	1,06291802	1,089929381	1,118008903
10		0	0,6	1,104431568	1,139155391	1,10716083	1,146048226
11		0	0,7	1,165214121	1,16525288	1,181893295	1,216899106
12		0	0,8	1,181725599	1,187485838	1,218323404	1,232730082
13		0	0,9	1,229567078	1,244361277	1,300435194	1,25567936
14		0	1	1,29739229	1,309477773	1,335482921	1,320790968
15		0,1	0	1,037194205	1,019586911	1,004268888	1,036835566
16		0,1	0,1	1,01324022	1,027804063	1,030623177	1,053817914

Source: Own construction

3.1 Quality of the selected model

First, the results of the Auto.SARIMA and the Auto.AIC procedures from the point of view of the quality criterion are presented. In the case of the time series generated by the ARIMA(1,0,1) or the SARIMA(1,0,1)(1,0,1)$_{12}$ models the conditions of "quality" for the non-seasonal parts are the following:

$$\left|\frac{\hat{\phi}_1 - \phi_1}{s(\hat{\phi}_1)}\right| < t_{0.975} \quad \text{and} \quad \left|\frac{\hat{\theta}_1 - \theta_1}{s(\hat{\theta}_1)}\right| < t_{0.975}, \tag{2}$$

where $t \sim t(T - 1)$, T is length of time series. If the selected models fulfill the above mentioned conditions they are denoted as “valid” models. The analogous criteria are applied to the seasonal parts of the models.

The results with the percentage quantifications are shown in Figure 1.

Figure 1 Quality Comparison of AIC, SARIMA

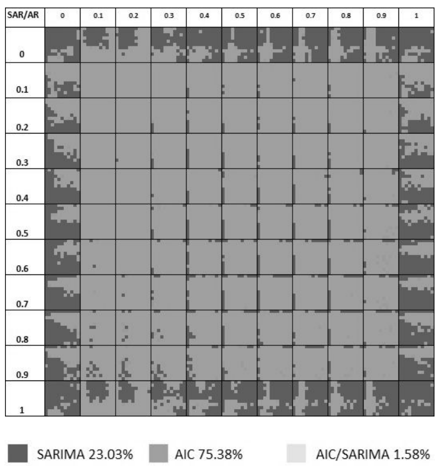
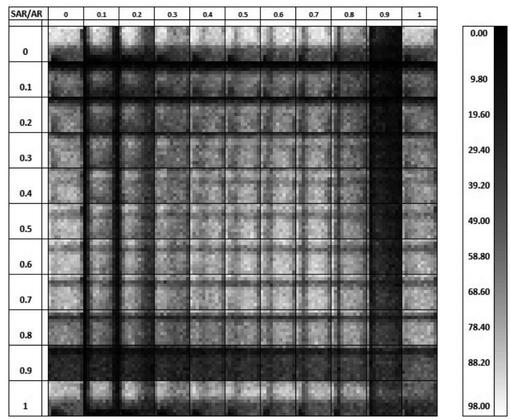


Figure 2 The Quality – Auto.SARIMA



Source: Authors’ calculations

It is obvious that the Auto.AIC is better than the Auto.SARIMA in 75.4% of cases. The Auto.SARIMA achieves better results in 23% of cases. Identical results are found in 1.6% of cases. But it is clear that there is a general group of the generating processes for which the Auto.SARIMA is better than the Auto. AIC. They are mainly the seasonal and the non-seasonal non-stationary (integrated) processes, and those processes that do not contain the non-seasonal and the seasonal autoregressive parts (AR respectively SAR). Furthermore, this procedure is superior to the processes that partly do not contain the nonseasonal and the seasonal moving average parts (MA, SMA). All these processes can be denoted as marginal. The results show that, mainly there, the “classical” model identification analysis represented by the Auto.SARIMA procedure (unit root testing, residual autocorrelation testing, normality and conditional heteroscedasticity testing and parameters estimate testing) has considerable importance.

Figure 2 shows the quality criterion (the percentage of the correct model selections) for the Auto. SARIMA procedure. It can be seen that this procedure has problems with the near nonseasonal and the near seasonal non-stationary processes; i. e., for the processes with the parameters $\phi_1 = 0.9$ and $\Phi_1 = 0.9$. In the first case, the success rate is 29%, and in the second it is 22.5%. The processes with the low values of the parameters; i. e., when the parameters ϕ_1 and Φ_1 lie between 0.1 and 0.2 together with the parameters ϕ_1, and Φ_1 between 0 and 0.2, while on the contrary, the seasonally non-stationary processes, when $\Phi_1 = 1$, create more problem areas. For the proceses with parameters ϕ_1 and Φ_1 between 0.3 and 0.7, the Auto.SARIMA gives good results regardless of the values of θ_1 and Θ_1. The success rate in this area is 66.1%. The average overall success rate of this procedure is 50.6%.

Figure 3 shows the quality criterion for the Auto.AIC procedure. Also, this procedure has problems with the near nonseasonal and the near seasonal non-stationary processes. In the case of $\phi_1 = 0.9$, the success rate is 37.8%, and when $\Phi_1 = 0.9$, the rate is 34%. The problematic areas are also for $\phi_1 = 0, 1$ and $\Phi_1 = 0, 1$, together with practically any values of parameters θ_1 and Θ_1. For the processes with parameters ϕ_1 and Φ_1 between 0.1 and 0.8, the Auto.AIC gives good results regardless of the values of θ_1 and Θ_1. The success rate in this area is 82.8%. The average overall success rate of this procedure is 66.7%. In comparison with the Auto.SARIMA, the Auto.AIC procedure is better.

Figure 3 The Quality – Auto.AIC

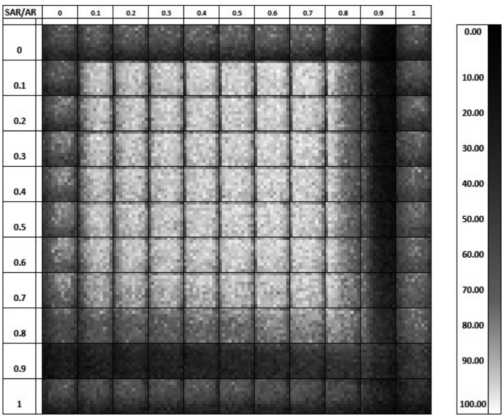
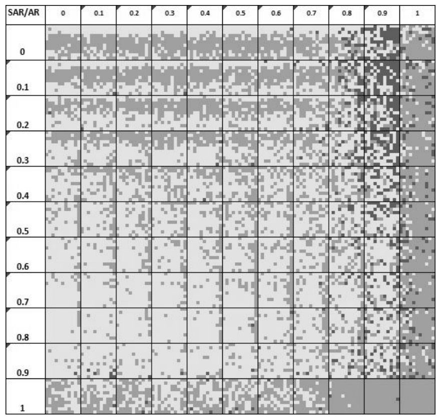


Figure 4 RMSE –1% tolerance



Source: Authors' calculations

3.2. Forecasts

The forecasts RMSE criterion is presented in the same way as the quality criterion. For each generating process, the procedure, which gives the the minimal value of the forecast RMSE, has been selected.

Figure 5 RMSE – Auto.AIC

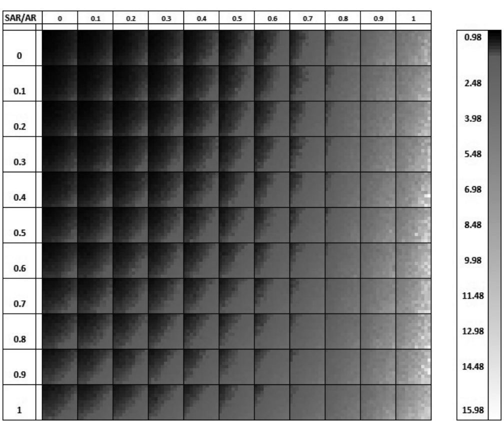
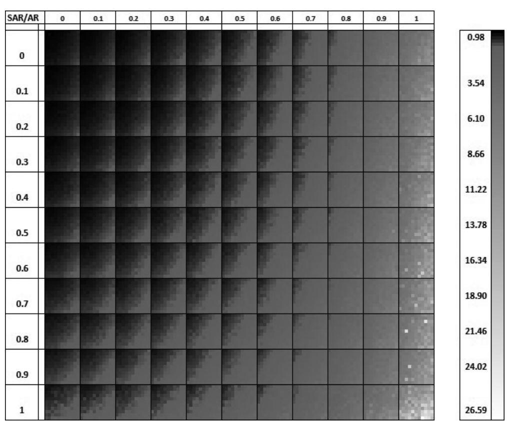


Figure 6 RMSE – Auto.SARIMA



Source: Authors' calculations

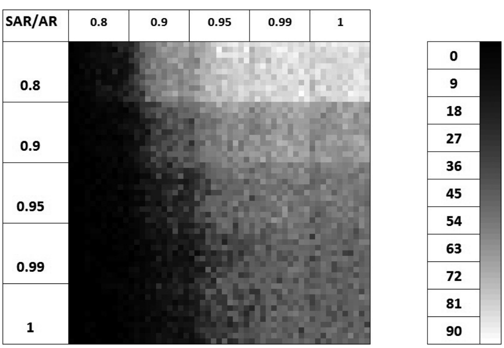
As the differences in the RMSE for the Auto.AIC and the Auto.SARIMA procedures are often very small, and the forecasts are very similar, it is suitable to compare them based on the tolerance limit of 1.0%. It means that the forecasts which are different in the RMSE up to 1.0% will be considered to be the same. Figure 4 ilustrates the results according which the Auto.SARIMA procedure gives the best forecasts in 5% of cases; the Auto.AIC in 36.7% cases. There are similar forecasts by both procedures in 58.3% cases. The Auto.AIC is better mainly for the non-seasonally non-stationary processes and the Auto.SARIMA for the near non-seasonally non-stationary processes.

Figure 5 shows the RMSE of the forecasts computed by the Auto.AIC procedure for the individual processes. It can be seen that along with the growing parameter values, the RMSE grows as well. The best results are either for the processes with zero or small values of the parameters. The worst results are for the nonseasonal non-stationary processes. It is interesting that the seasonal nonstationarity does not have such a strong influence on the forecasts RMSE like the nonseasonal nonstationarity. Figure 6 shows the RMSE of the forecasts computed by the Auto.SARIMA. The pattern is similar to that in Figure 5.

3.3 Forecasting of the nearly integrated time series

In this part we will extend the above analysis about the situation of so called near integrated, but still stationary processes. Figure 7 depicts the forecasting success of the nonseasonal integrated model of the SARIMA(0,1,1)(1,0,1)$_{12}$ type for this type of process, irrespective of the forecasting procedure. It can be seen that even for the non-seasonally stationary processes with ϕ_1 from 0.90 to 0.95, the integrated model is more suitable for forecasting than the correct stationary model. This result is consistent with the result for the example of Pincheira and Medel (2016). The possible explanation is that the estimates of the parameters of the correct models for the time series generated by the nearly non-stationary processes have greater variability and are thus less accurate.

Figure 7 The Forecasting Success of SARIMA(0,1,1)(1,0,1)



Source: Authors' calculations

CONCLUSION

The goal of the simulation study was to analyze the relationship of the size of the systematic part of the process (it is given by the magnitude of the SARIMA parameters, bigger values of parameters mean stronger systematic part), which is used for time series generation and the ability to select the proper model for the generated time series. The second goal was to analyze the quality of forecasts for the time series generated by the processes with different systematic parts. In this connection the analysis of the ability to select suitable model and reach the relatively accurate forecasts for the time series generated by the near non-stationary and the non-stationary processes was also important.

As a results of the simulation study, the following facts have been found:

1. The Auto.AIC procedure is better for the selection of models for the time series generated by the stationary and invertible processes. The Auto.SARIMA procedure is better for the modelling the time series from so called marginal processes; i. e. mainly from the non-stationary processes and the processes that do not contain the non-seasonal and the seasonal autoregressive parts.
2. For both procedures it is difficult to find the correct model for the time series generated by processes with low values of the autoregressive parameters, and by the near non-stationary processes.

In the first case, the systematic part of the time series is very weak and the property which we are looking for does not show sufficient transparency, so it is possible to overlook it. In the second case, the two different and incompatible situations have the same, or very similar effects, so it is difficult to distinguish between them.

3. The Auto.AIC procedure leads to the better forecasts, but for near to non-stationary processes the Auto.SARIMA procedure is better. The differences in the accuracy between the Auto.SARIMA and Auto.AIC procedures are relatively small. With the growing magnitude of parameters, the accuracy of forecasts decreases in the case of both procedures.
4. For the forecasting of the time series generated by the non-seasonally nearly integrated processes, the non-seasonally integrated models are more suitable than the correct stationary ones.

ACKNOWLEDGEMENTS

This paper was written with the support of the Czech Science Foundation project No. 402/12/G097 DYME – *Dynamic Models in Economics.*

References

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, 19(6), pp. 716–723.

BOX, G. E. P, JENKINS, G. M., REINSEL, G. C., LJUNG, G. M. *Time Series Analysis: Forecasting and Control.* Wiley, 2015.

BROCKWELL, P. J. AND DAVIS, R. A. *Introduction to Time Series and Forecasting.* Springer, 2010.

CANOVA, F. AND HANSEN, B. E. Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability. *Journal of Business and Economic Statistics*, 1995, 13, pp. 237–252.

DICKEY, D. A. AND FULLER, W. A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Stat. Association*, 1979, 74, pp. 427–431.

ENDERS, W. *Applied Econometric Time Series.* Wiley, 2014.

ENGLE, R. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 1982, 50, pp. 987–1008.

HAMILTON, J. D. *Time Series Analysis.* Princeton University Press, 1994.

HANNAN, E. J. AND QUINN, B. G. The Determination of the Order of an Autoregression. *Journal of Royal Statistical Society*, 1978, 41, pp. 190–195.

JARQUE, C. AND BERA, A. Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals. *Economics Letters,* 1980, 6, pp. 255–259.

KWIATOVSKI, D., PHILLIPS, P. C. B., SCHMIDT, P., SHIN, Y. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics*, 1992, 54, pp. 159–178.

LJUNG, G. M. AND BOX, G. E. P. On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 1978, 65, pp. 297–303.

PESARAN, M. H. *Time Series and Panel Data Econometrics.* Oxford University Press, 2015.

PHILLIPS, P. C. B. AND PERRON, P. Testing for a Unit Root in Time Series Regression. *Biometrika*, 1988, 75, pp. 335–346.

PINCHEIRA, P. M. AND MEDEL, C. A. Forecasting with a Random Walk. *Finance a úvěr – Czech Journal of Economics and Finance*, 2016, 66(6), pp. 539–564.

R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing* [online]. R Foundation for Statistical Computing, Vienna, Austria, 2008. <*http://www.R-project.org*>.

SCHWARTZ, G. Estimating the Dimension of a Model. *Annals of Statistics*, 1978, 6, pp. 461–464.

TRCKA, P. *Výstavba lineárnych stochastických modelov časových radov triedy SARIMA – automatizovaný postup.* Dipl. thesis, University of Economics, Prague, 2015.

WEI, W. W. S. *Time Series Analysis: Univariate and Multivariate Methods.* 2nd Ed. Pearson Education, 2006.