FORECASTING THE INVENTORY LEVEL OF MAGNETIC CARDS IN TOLLING SYSTEM

Bratislav Lazić [a], Nebojša Bojović [b], Gordana Radivojević [b*], Gorana Šormaz [a]

[a] University of Belgrade, Mihajlo Pupin Institute, Serbia
[b] University of Belgrade, Faculty of Transport and Traffic Engineering, Serbia

Abstract: *Forecasting the inventory level of magnetic cards is an important prerequisite for the functioning of a tolling system. This paper presents the approach to forecasting the required number of magnetic cards in the tolling system on the Belgrade – Niš highway section in Serbia, as a prerequisite for purchasing and forming an optimal inventory level. The* $ARIMA(1,1,2)\times(1,1,0)_{12}$ *model was developed and applied for forecasting the monthly inventory level of magnetic cards needed for the years 2015 and 2016.*

Keywords: *time series, forecasting, magnetic cards inventories, tolling system.*

1. INTRODUCTION

The tolling system is a complex system which enables the highway owner toll charging on various sections, and fast and high-quality usage of highways to the users. Toll charging in the Republic of Serbia is within the competence of Public Enterprise "Roads of Serbia", which is responsible for organization of work, supervision, management and technical provision for functioning of the system.

Toll charging is carried out on three sections of highways in the Republic of Serbia: Belgrade – Niš, Belgrade – Šid and Belgrade – Subotica. The Belgrade – Šid and Belgrade – Niš sections have a closed system of toll charging in which the toll amount depends on vehicle category and travel distance (PE "Roads of Serbia", 2015). A highway user gets a magnetic card at entrance toll plazas and pays toll at the exit plaza, after card reading. The users get the magnetic cards at entrance plazas from distributors (automatic regime) or pre-magnetized cards from operators (manual regime). At collecting plazas, the cards are put aside, namely put in storages until they are destroyed. The entrance plazas must have an adequate level of card inventories to enable undisturbed functioning of vehicle entrance and traffic on highway.

PE "Roads of Serbia" deals with the organization of toll collection and accordingly has to plan the following: magnetic card acquisition, making sufficient amount of pre-magnetized cards, organization of workers at collection stations, the number of entrance/exit lanes at plazas as well as regular maintenance of toll collection system. It is necessary to precisely forecast the number of magnetic cards in the toll collection system for these plans to meet the future

[*] g.radivojevic@sf.bg.ac.rs; gordana.radivojevic@pupin.rs

requirements of the system. The forecast is based on the number of issued magnetic cards in the previous time periods, namely in analyzing the monthly time series of the number of issued magnetic cards.

A Multiplicative Seasonal ARIMA class of non-stationary time series was applied in this paper for the analysis of time series. The used time series includes data on the number of magnetic cards from January 2000 till November 2014 on the Belgrade – Niš highway section. The obtained model was applied for forecasting the required number of magnetic cards for the period December 2014 – December 2016.

The paper has been organized in the following way. The second part describes the theoretical basis and Box-Jenkins methodology for analyzing and modelling time series. The third part shows application of the Box-Jenkins methodology for modelling the inventory level of magnetic cards for toll collection system on the Belgrade – Niš highway section. The conclusion and possible directions of future research are given in the fourth part.

2. TIME SERIES ANALYSIS

Time series can be described as a set of temporal ordered realizations of a random variable during a series of successive time periods. By analyzing changes in the random variable in time, it is sometimes possible to define a performance model which enables forecasting future states.

Seasonal time series consist of periodical changes which repeat in certain time intervals. The smallest repetitive time period is called seasonal period (s). When realization of a random variable repeats after seasonal period with certain regularities, it may be expected that the values in seasonal periods will correlate mutually. High values of autocorrelation function on seasonal realizations are indicators of seasonal non-stationary process. Seasonal difference operator$(1 - B^S)$ is used eliminating seasonal non-stationarity.

By combining seasonal difference operator with ARIMA model, Box and Jenkins defined the Multiplicative Seasonal ARIMA model (Box et al., 1994; Brockwell, Davis, 2002; Cryer, Chan, 2008) as:

$$(1-B)^d(1-B^s)^D \varphi_p(B)\Phi_P(B^s)X_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \qquad (1)$$

In this relation B is lag operator, $\varphi_p(B)\Phi_P$ and $\theta_q(B)$ are autoregressive polynomial and moving average polynomial, and $\Phi_P(B^s)$ i $\Theta_Q(B^s)$ are seasonal autoregressive polynomial and seasonal moving average polynomial. Multiplicative Seasonal ARIMA model with seasonal period s is marked as ARIMA$(p, d, q) \times (P, D, Q)_s$, where p and q are orders of non-seasonal autoregressive polynomial and seasonal moving average polynomial, P and Q are orders of seasonal autoregressive polynomial and moving average polynomial, d is non-seasonal and D is seasonal differencing degree (Box et al., 1994; Wei, 2006; Cryer, Chan, 2008).

Box and Jenkins proposed a methodology for analysing and modelling time series which includes four stages (Box et al., 1994):

- Model Identification,
- Model Estimation,
- Model Validation, and
- Forecasting.

Model Identification consists of a series of procedures over time series data, in order to chose the corresponding model from ARIMA model set. These procedures include the following (Tsui et al., 2014):

- Time series plotting – enables detecting trend occurrence and seasonal changes in time series. If it is noted that mean value is not constant then logarithmic data transformation is applied.
- Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) – based on theoretical characteristics of ACF and PACF functions for ARIMA models, p and q orders of ARIMA model are assumed. These diagrams can indicate seasonal changes in time series.
- Unit root tests (Augmented Diskey-Fuller test, Philips-Perron test) – tests are carried out by changing length of lag operator until stationary series is obtained. Existence of unit roots indicates to the necessity of series differencing.
- Seasonal HEGY test – if there are seasonal unit roots then seasonal differencing is applied.

The described procedures give the set of values of ARIMA model (p, q, d, P, Q, D) parameters.

Model Estimation implies that coefficients of autoregressive polynomial and moving average polynomial are determined for every possible parameter (p, q, d, P, Q, D) namely fitting of model is in question. In this phase the methods of non-linear least squares and maximum likelihood estimation are applied. The best model is the one which has the lowest information criterion, which may be (Kirchgässner and Wolters, 2007; Cryer, Chan, 2008):

- Akaike information criteria: $AIC = T\ \ln(MSE) + 2k$
- Baesian information criteria: $BIC = T\ \ln(MSE) + k\ \ln(T)$
- Normalized Baesian information criteria: $Normalized\ BIC = \ln(MSE) + \frac{k\ \ln(T)}{T}$
- Hanan-Quin information criteria: $HQIC = T\ \ln(MSE) + 2k\ln\big(\ln(T)\big)$

Model Validation involves checking residuals – the differences of realized and forecasted values. ACF and PACF residual functions are used for this or some of the formal approaches (eg. Ljung-Box Q-statistics). If there is no autocorrelation in residuals, the model is adequate. On the contrary, model is inadequate and it is necessary to go back to identification stage.

Forecasting implies using the obtained model to forecast values for h step ahead. By comparing the obtained values with h past values of series realization, which were not used while estimating model parameters, it can be evaluated whether the chosen model has good quality forecast. Model estimation can be done according to any of the following criteria:

- Mean absolute deviation $MAD = \frac{1}{n}\sum_i \left|X_i - \hat{X}_i\right|$
- Mean absolute percentage error $MAPE = \frac{100}{n}\sum_i \left|\frac{X_i - \hat{X}_i}{X_i}\right|$
- Mean square error $MSE = \frac{1}{n}\sum_i \left(X_i - \hat{X}_i\right)^2$
- Root mean square error $RMSE = \sqrt{MSE}$

3. ARIMA MODEL FOR FORECASTING NUMBER OF MAGNETIC CARDS

In order to forecast the necessary inventory level of magnetic cards for toll collection on the Belgrade – Niš section, the time series of total number of issued magnetic cards was used from the period January 2000 – November 2014 (PE "Roads of Serbia", 2000-2014). The time series consists of 179 monthly realizations, out of which the first 168 were used for estimating model parameters and the last 11 for model performance estimation.

3.1. Identification of time series model

Time series plotting of issued magnetic cards is shown in Figure 1. An analysis of the time series plot shows non-stationary time series with an increasing trend and seasonal variation. Seasonal variations are justified by the fact that the largest number of highway users is in the period July-August during the holidays, with heavy transit of traffic from Western European countries towards Turkey, Bulgaria and Greece, and vice versa, and that the lowest traffic on the highway occurs during winter weather conditions, especially in months with a great number of exceptionally cold days and snowfall.

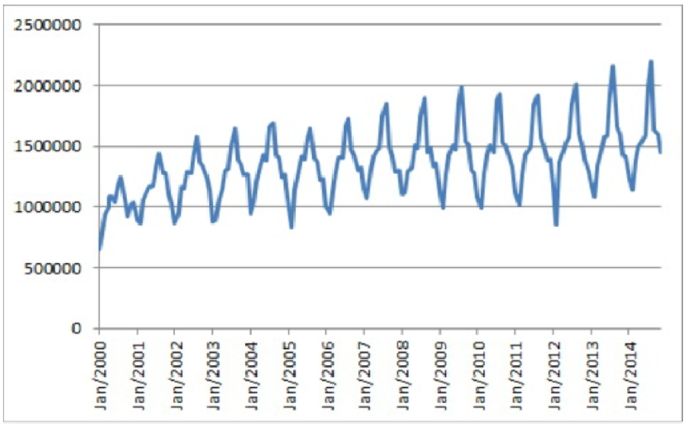
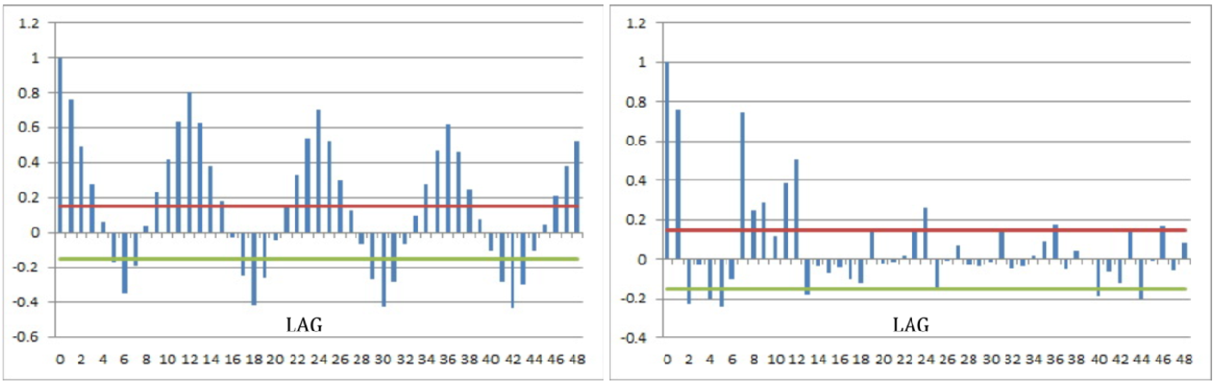


Figure 1. Time series of total number of magnetic cards for the period 2000–2014.

In order to stabilize variance, a logarithmic time series transformation was carried out. The transformed series in this paper is titled ln time series. An autocorrelation function (ACF) and partial autocorrelation function (PACF) were formed for ln time series, shown in Figures 2(a) and 2(b).



(a) ACF (b) PACF

Figure 2. Autocorrelation and partial autocorrelation function ln time series

Great correlation function values were marked in all time periods. Gradual decrease of autocorrelation function values indicate to the existence of a long-term trend, and points out to the fact that ln time series differencing is necessary to be carried out when making model, namely that $d = 1$ in ARIMA model. High autocorrelation function values on annual periods indicate to the need for seasonal differencing on model. By applying differencing operator $(1 - B)(1 - B^{12})$ stationary time series was obtained. In the model identification phase, formal tests for unit roots presence were carried out. The results of Dickey-Fulle test and Philips-Person test confirmed stationarity.

3.2. Model Estimation

Model identification showed that ARIMA$(p, 1, q) \times (P, 1, Q)_{12}$ models proved to be most suitable for the observed ln time series. In the model estimation stage, all possible models from alternative model set, where $p, q = 0, 1, 2$ and $P, Q = 0, 1$, were considered. By using MATLAB Econometrics Toolbox®, the autoregressive polynomial and moving average polynomial coefficient were estimated as well as seasonal autoregressive polynomial and moving average polynomial coefficient for all observed models. For each of these models, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values were found and Ljung-Box Q-test carried out. ARIMA$(1, 1, 2) \times (1, 1, 0)_{12}$ model represents the best model for ln time series, because the lowest AIC and BIC criterion values was obtained. The parameter values of this model for ln time series are shown in Table 1, and the model relation is given by equation (2):

$$(1-B)(1-B^{12})(1+0.354878B)(1+0.715782B^{12})X_t = (1+0.109707B+0.416844B^2)\varepsilon_t \qquad (2)$$

Table 1. ARIMA model parameters values for ln time serie

Parameter	Value	Standard deviation	*t*-statistics
AR{1}	-0.354878	0.210124	-1.6889
SAR{12}	-0.715782	0.0409527	-17.4783
MA{1}	-0.109707	0.210714	-0.52065
MA{2}	-0.416844	0.134475	-3.09978
Variance	0.00240232	0.000191872	12.5204
AIC	-522.4981		
BIC	-500.6303		
MAPE	0.2614		
RMSE	0.0579		
R-squared	0.9433		

3.3. Model Validation

Diagnostic checking confirms that the chosen model is stationary and that there are no redundant parameters. Residuals analysis (Ljung-Box Q=11.6542, p-value 0.92747>0.05) indicate residuals are strictly random and that no significant autocorrelation exists between residuals on different time periods.

3.4. Magnetic Cards inventory level forecasting

The differences between anticipated values obtained by model and realized values of time series for the period January-November 2014 are given in Figure 3(a).

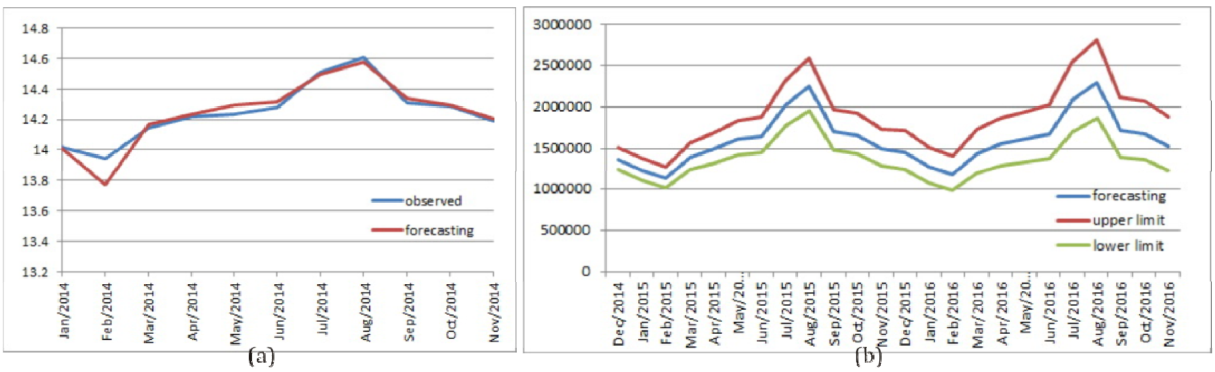


Figure 3. Inventory level of the magnetic cards forecast

As the mean absolute percentage error (MAPE) and root mean square error (RMSE) values are small (MAPE $\leq$ 10%) (Tsui et al., 2014), it may be concluded that the obtained model is good enough for forecasting the inventory level of magnetic cards on the Belgrade – Niš highway section.

By using the obtained model, the number of magnetic cards for the period December 2014 – November 2016 was forecasted. Forecast value and 95% confidence interval are given in Figure 3(b).

4. CONCLUSION

The aim of this paper was to forecast inventory level of magnetic cards needed to serve demand of PE "Roads of Serbia". High quality forecasting enables the company to properly plan acquisition of magnetic cards, the making and distribution of pre-magnetized cards, the required personnel and organize regular maintenance of toll charging system. The Box-Jenkins methodology was applied for the choice of the adequate ARIMA model based on time series of monthly number of issued cards. The obtained $\mathrm{ARIMA}(1,1,2) \times (1,1,0)_{12}$ model for the Belgrade-Niš highway section showed high performance and may be used for high-quality forecasting of the magnetic cards inventory level. Forecasting of the required number of magnetic cards for the years 2015 and 2016 was effected by applying model. For further improvement of the model, it is necessary to also include other variables which affect traffic intensity on highway (gross national product, real wages rate, gas prices, etc.).

ACKNOWLEDGMENT

This work was supported by the Ministry of Education and Science of the Government of the Republic of Serbia through the projects TR 36005 and TR36022, for the period 2011-2015.

REFERENCES

[1] Box G., Jenkins G., Reinsel G. (1994). Time Series Analysis Forecasting and Control. Third Edition. Prentice Hall.

[2] Brockwell P., Davis R. (2002). Introduction to Time Series and Forecasting. Second Edition. Springer.

[3] Cryer J.D., Chan K. (2008). Time Series Analysis With Applications in R. Springer.

[4] PE "Roads of Serbia" (2000-2014). Publication counting traffic on public roads of the Republic of Serbia.

[5] PE "Roads of Serbia" (2015). www.putevi-srbije.rs. (March, 2015).

[6] Kirchgässner G., Wolters G. (2007). Introduction To Modern Time Series Analysis. Springer.

[7] Matlab Statistics Toolbox User's Guide R2014b. (2014). MathWorks.

[8] Tsui W.H.K, Baili H.O., Gilbery A., Gow H. (2014). Forecasting of Hong Kong airport's passenger throughput, Tourism Management 42, 62-76.

[9] Wei W. (2006). Time Series Analysis: Univariate and Multivariate Methods. Pearson Addison Wesley.