

## FORECASTING TOURISM DEMAND IN CROATIA USING BOX AND JENKINS METHODOLOGY

**Damira Đukec**

<https://doi.org/10.20867/tosee.05.18>

### **Abstract**

Tourism is a significant sector in Croatian economy, generating income and employment. Recent data show that contribution of tourism to GDP is almost 20%. Planning and forecasting is crucial for further development. Demand forecasting models are therefore essential input in developing a competitive tourism industry which enables sustainable growth. Tourism demand forecasting methods can be divided in qualitative and quantitative methods. Econometric models differ from time series in identifying the casual relationships between variables. Time series models mostly rely on Box and Jenkins autoregressive integrated moving average (ARIMA) or seasonal ARIMA (SARIMA) methodology. For the purpose of this paper we will focus on quantitative methods. Quantitative methods used for forecasting purposes are either time series models or econometric studies. In this paper tourism demand for Croatia has been modelled using Box and Jenkins seasonal ARIMA methodology. The Box-Jenkins methodology refers to a set of procedures for identifying and estimating time series models within the class of autoregressive integrated moving average (ARIMA) models. The data for tourist arrivals showed clear patterns of seasonality and an upward trend. After differentiation and seasonal adjustment, a model selection and diagnostic checking followed. The model selected was seasonal ARIMA (2,1,1) (0,1,1)<sub>12</sub>. The model fits the observed data well and shows no autocorrelation of the residuals.

**Keywords** Tourism demand, Croatia, forecasting, Box and Jenkins, ARIMA

### **INTRODUCTION**

Tourism is a significant sector in Croatian economy, generating income and employment. Recent data show that contribution of tourism to GDP is almost 20%. Planning and forecasting is crucial for further development. Demand forecasting models are therefore essential input in developing a competitive tourism industry which enables sustainable growth. In this paper a forecasting model will be created for tourism demand in Croatia, using Box and Jenkins methodology. The model will enable more accurate forecasts of tourism demand which will lead to improvement in strategic planning.

Tourism demand forecasting methods can be divided in qualitative and quantitative methods. For the purpose of this paper we will focus on quantitative methods. Quantitative methods used for forecasting purposes are either time series models or econometric studies. The methodologies applied in both of these categories of models, will be further explored discussing their strengths and weaknesses. Econometric models differ from time series in identifying the casual relationships between variables. Time series models mostly rely on Box and Jenkins autoregressive integrated moving average (ARIMA) or seasonal ARIMA (SARIMA) methodology. Generalised Autoregressive

Conditional Heteroscedastic (GARCH) models are also used as an extension of univariate time series analysis. Econometric models use explanatory variables such as tourist income, tourism prices in a destination relative to origin country, tourism prices of competing destinations and exchange rates to model and predict tourism demand. Some of the techniques applied are regression analysis based on ordinary least squares (OLS), error correction models (ECM), vector autoregressive models (VAR), time varying parameter (TVP), structural equation modelling (SEM), autoregressive distributed lagged model (ADLM) and the almost ideal demand system (AIDS). Other than above mentioned methods, recent developments include empirical applications of artificial intelligence (AI) such as artificial neural network method (ANN), the fuzzy time series method and genetic algorithms (GAs).

This paper consists of five chapters. After introduction there is a short review of recent literature in the field of tourism demand forecasting. In the third chapter theoretical background of the forecasting methodology in this paper is described. Results of the tourism demand forecasting modelling are presented in the fourth chapter. The paper ends with a conclusion where the main results of the research are summarised.

## 1. LITERATURE REVIEW

Tourism demand forecasting is an area of interest for many researchers. In the table below there is only a fraction of papers in the subject that indicate versatility of research methods and applications. There are few examples of modelling tourism demand in Croatia but none of them using Box and Jenkins methodology for modelling total (both domestic and foreign tourist arrivals as a measure of demand) tourism demand. A thorough review of literature undertaken by Song and Li (2008) confirms the conclusion of earlier research that there is no evidence of one model consistently outperforming other models. Some recent developments in the field include the use of AI techniques and the use of forecast combination and forecast integration of quantitative and qualitative approaches which leads to improvement in forecast accuracy.

Table 1: **Recent empirical applications**

Author	Methodology	Country
Petrevska (2017)	ARIMA	Macedonia
Tica and Kožić (2015)	Leading indicator	Croatia
Cerović, Grudić Kvasić and Ivančić (2017)	Exponential smoothing	Croatia
Song and Witt (2006)	VAR	Macau
Chhorn and Chaiboonsri (2017)	GARCH	Cambodia
Lee (2010)	ECM	Hong Kong
Song, Li, Witt and Athanasopoulos (2011)	TVP	Hong Kong
Baldigara and Koić (2015)	OLS	Croatia
Constantino et al. (2016)	ANN	Mozambique
Turner and Witt (2001)	SEM	New Zealand
Han et al. (2006)	AIDS	Europe

Author	Methodology	Country
Song et al. (2003)	ADLM	Hong Kong
Sakhujia et al. (2016)	GA Fuzzy Time Series	Taiwan
Singh (2013)	SARIMA	Bhutan

Source: author

## 2. THEORETICAL BACKGROUND

In this paper tourism demand for Croatia will be modelled and forecasted using Box and Jenkins methodology. The Box-Jenkins methodology refers to a set of procedures for identifying and estimating time series models within the class of autoregressive integrated moving average (ARIMA) models. The Box-Jenkins method refers to the iterative application of the following three steps: 1. Identification. Using plots of the data, autocorrelations, partial autocorrelations, and other information, a class of simple ARIMA models is selected. This amounts to estimating appropriate values for  $p^1$ ,  $d^2$ , and  $q^3$ . 2. Estimation. The phis and thetas of the selected model are estimated using maximum likelihood techniques, backcasting, etc., as outlined in Box-Jenkins (1976). 3. Diagnostic Checking. The fitted model is checked for inadequacies by considering the autocorrelations of the residual series. Monthly data on tourist arrivals, domestic and total, were used as a measure for tourism demand. The data showed a clear seasonal pattern so a seasonal ARIMA was used for modelling.

To deal with series containing seasonal fluctuations, Box-Jenkins recommend the following general model:

$$\phi_p(B) \varphi_P(B^s)(1 - B)^d(1 - B^s)^D Y_t = \theta_q(B) \Theta_Q(B^s) \varepsilon_t \quad (1)$$

where  $d$  is the order of differencing,  $s$  is the number of seasons per year, and  $D$  is the order of seasonal differencing. The operator polynomials are

$$\begin{aligned} \phi_p(B) &= (1 - \phi_1 B - \dots - \phi_p B^p) \\ \theta_q(B) &= (1 - \theta_1 B - \dots - \theta_q B^q) \\ \varphi_P(B^s) &= (1 - \varphi_1 B - \dots - \varphi_P B^{sP}) \\ \Theta_Q(B^s) &= (1 - \Theta_1 B - \dots - \Theta_Q B^{sQ}) \end{aligned}$$

Note that  $(1 - B^s) Y_t = Y_t - Y_{t-s}$ .

Box-Jenkins explain that the maximum value of  $d$ ,  $D$ ,  $p$ ,  $q$ ,  $P$ , and  $Q$  is two. Hence, these operator polynomials are usually simple expressions.

<sup>1</sup>  $p$  is the order of the autoregressive process i.e. the number of AR terms

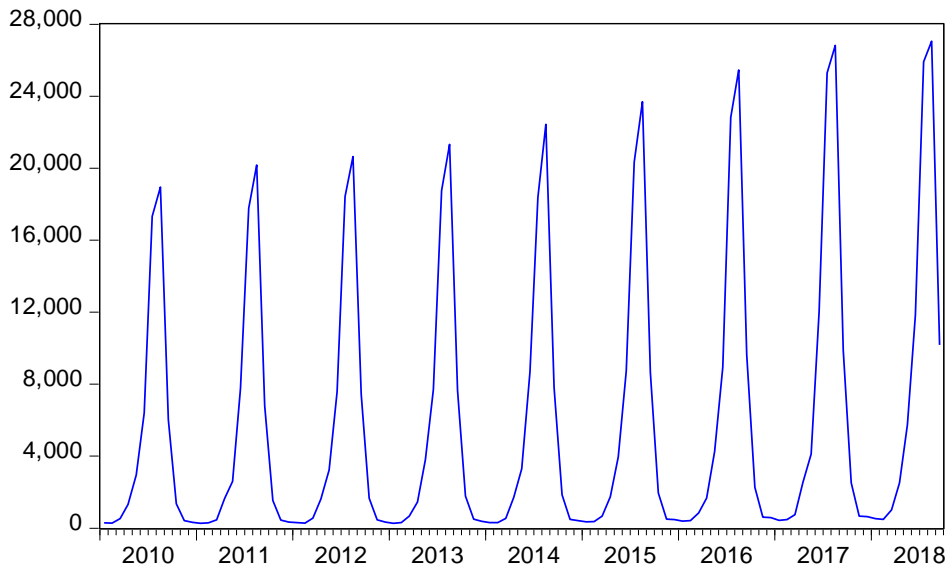
<sup>2</sup>  $d$  shows the number of times the series has to be differenced before it becomes stationary

<sup>3</sup>  $q$  is the order of the moving average process i.e. the number of MA terms

### 3. EMPIRICAL RESULTS

In this section the results of the empirical research will be presented. To find a forecasting model for Croatian tourism demand Box and Jenkins approach was used. The data used in this research are monthly tourist arrivals for period from January 2010 to September 2018 and are downloaded from the web site of Croatian Bureau of Statistics. The observed time period makes 105 observations which is considered to be a large sample for seasonal forecasting (Hyndman and Kostenko, 2007). As we can see from figure 1. The data show a clear seasonal pattern with a growth trend. Also, we can detect a presence of heteroscedasticity in the data.

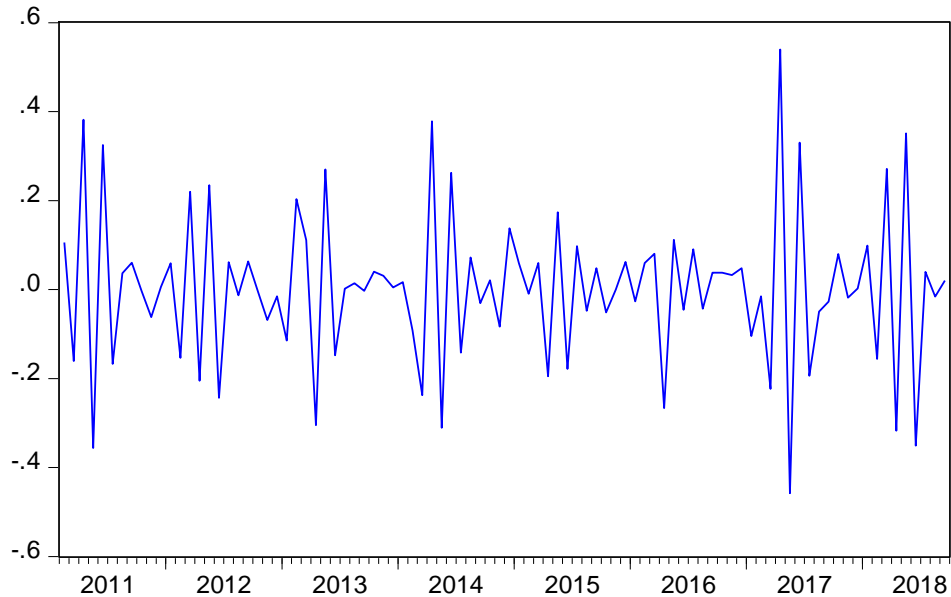
Figure 1: **Monthly tourist arrivals to Croatia**



Source: output of statistical software EViews 9

Further modelling requires transformation of the data described above. To remove the growth trend and seasonality the data has been differenced. Also, a log transformation of the data has been made to remove the heteroscedasticity. After transformation the data (figure 2) appears to be stationary. To test the stationarity Augmented Dickey Fuller test has been applied (table 2).

Figure 2: **Monthly tourist arrivals to Croatia – transformed data**



Source: output of statistical software EViews 9

The results of the ADF test are shown in the table below. The null hypothesis is that transformed data have a unit root, which would imply that the data is nonstationary. With the probability of 0,0000 the null hypothesis is rejected leading to a conclusion that the data are stationary. After achieving stationarity, the next step in the modelling process is to select a model.

Table 2: **Augmented Dickey-Fuller test**

Null Hypothesis: DTOTAL1 has a unit root  
 Exogenous: None  
 Lag Length: 1 (Automatic - based on SIC, maxlag=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-12.25502	0.0000
Test critical values:		
1% level	-2.590910	
5% level	-1.944445	
10% level	-1.614392	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
Dependent Variable: D(DTOTAL1)  
Method: Least Squares  
Date: 01/24/19 Time: 14:04  
Sample (adjusted): 2011M04 2018M09  
Included observations: 90 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DTOTAL1(-1)	-2.341398	0.191056	-12.25502	0.0000
D(DTOTAL1(-1))	0.309198	0.100837	3.066300	0.0029
R-squared	0.905851	Mean dependent var		0.002010
Adjusted R-squared	0.904781	S.D. dependent var		0.333875
S.E. of regression	0.103026	Akaike info criterion		-1.685699
Sum squared resid	0.934064	Schwarz criterion		-1.630148
Log likelihood	77.85645	Hannan-Quinn criter.		-1.663297
Durbin-Watson stat	1.946123			

Source: output of statistical software EViews 9

The model selection starts with the detail examination of autocorrelogram and partial autocorrelogram of the data. As we can see both ACF and PACF slowly decay after first few significant spikes which indicates a possible mixed model with both AR and MA processes present. Also we detect significant spikes around lag 12 in ACF implying seasonal MA process.

Figure 3: **ACF and PACF**

Date: 01/24/19 Time: 13:59  
Sample: 2010M01 2019M09  
Included observations: 92

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
***** .	***** .	1	-0.790	-0.790	59.293	0.000
. **** .	** .	2	0.506	-0.314	83.891	0.000
** .	. .	3	-0.299	-0.065	92.599	0.000
. * .	* .	4	0.114	-0.183	93.869	0.000
. .	** .	5	-0.044	-0.219	94.065	0.000
. .	* .	6	0.021	-0.128	94.108	0.000
. .	. .	7	0.041	0.073	94.276	0.000
* .	* .	8	-0.123	-0.152	95.830	0.000
. ** .	. .	9	0.216	0.070	100.69	0.000
** .	* .	10	-0.316	-0.153	111.20	0.000
. *** .	. * .	11	0.417	0.192	129.76	0.000
*** .	. .	12	-0.465	-0.052	153.17	0.000
. *** .	* .	13	0.379	-0.195	168.86	0.000
** .	. .	14	-0.220	0.063	174.24	0.000

Date: 01/24/19 Time: 13:59  
Sample: 2010M01 2019M09  
Included observations: 92

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.  *	.  .	15	0.089	0.036	175.13	0.000
.  .	. * .	16	-0.025	-0.120	175.20	0.000
.  .	.  .	17	0.027	0.011	175.28	0.000
.  .	.  .	18	-0.029	0.064	175.38	0.000
.  .	.  .	19	0.003	0.036	175.38	0.000
.  .	. * .	20	0.004	-0.191	175.38	0.000
.  .	.  **	21	0.033	0.216	175.51	0.000
. * .	. ** .	22	-0.110	-0.215	177.01	0.000
.  *	.  **	23	0.206	0.222	182.34	0.000
. ** .	. * .	24	-0.252	-0.094	190.44	0.000
.  *	. * .	25	0.205	-0.160	195.84	0.000
. * .	.  .	26	-0.151	-0.015	198.85	0.000
.  *	.  .	27	0.122	0.045	200.84	0.000
.  .	.  *	28	-0.046	0.091	201.13	0.000
.  .	.  .	29	-0.002	0.072	201.13	0.000
.  .	.  .	30	0.010	-0.021	201.14	0.000
.  .	.  .	31	-0.037	0.054	201.33	0.000
.  .	. * .	32	0.066	-0.079	201.95	0.000
. * .	. * .	33	-0.166	-0.113	205.98	0.000
.  **	. * .	34	0.311	-0.097	220.40	0.000
. *** .	. * .	35	-0.441	-0.079	249.87	0.000
.  ****	.  .	36	0.509	0.044	289.90	0.000

Source: output of statistical software EViews 9

Based on the initial analysis of ACF and PACF several models, following parsimony principle, have been estimated and evaluated. Finally, a model was selected based on the lowest Akaike information criterion. The selected model is seasonal ARIMA (2,1,1) (0,1,1)<sub>12</sub> can be written in its general form as follows:

$$(1 - \phi_2 B)(1 - B)(1 - B^{12})Y_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})\varepsilon_t \quad (1)$$

where,  $Y_t$  represents tourist arrivals,  $B$  is the backshift operator, and  $\varepsilon_t$  is white noise. Estimation results of the model (1) are shown in the table 3.

**Table 3: Model estimation results**

Dependent Variable: DLOG(TOTAL,1,12)  
 Method: ARMA Maximum Likelihood (OPG - BHHH)  
 Date: 01/24/19 Time: 13:39  
 Sample: 2011M02 2018M09  
 Included observations: 92  
 Convergence achieved after 22 iterations  
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.381253	0.114422	3.331989	0.0013
MA(1)	-0.928624	0.050090	-18.53920	0.0000
SMA(12)	-0.593661	0.099687	-5.955232	0.0000
SIGMASQ	0.007179	0.001100	6.526947	0.0000
R-squared	0.763653	Mean dependent var		0.001418
Adjusted R-squared	0.755595	S.D. dependent var		0.175238
S.E. of regression	0.086633	Akaike info criterion		-1.932817
Sum squared resid	0.660460	Schwarz criterion		-1.823174
Log likelihood	92.90958	Hannan-Quinn criter.		-1.888564
Durbin-Watson stat	2.140627			
Inverted AR Roots	.62	-.62		
Inverted MA Roots	.96	.93	.83-.48i	.83+.48i
	.48+.83i	.48-.83i	.00-.96i	-.00+.96i
	-.48+.83i	-.48-.83i	-.83+.48i	-.83-.48i
	-.96			

Source: output of statistical software EViews 9

The estimated seasonal ARIMA model can be written as follows:

$$(1 - 0,381253B)(1 - B)(1 - B^{12})Y_t = (1 + 0,928624_1B)(1 + 0,593661_1B^{12})\varepsilon_t \quad (2)$$

All the estimated coefficients are significant and  $R^2$  is 0,76 which indicates that the model fits the data well. Further analysis of the estimated model requires diagnostic checking of the residuals. The ACF and PACF figure below shows no indication of autocorrelation in the residuals.



Figure 4: **ACF and PACF of residuals**

Date: 01/24/19 Time: 14:14

Sample: 2010M01 2018M09

Included observations: 92

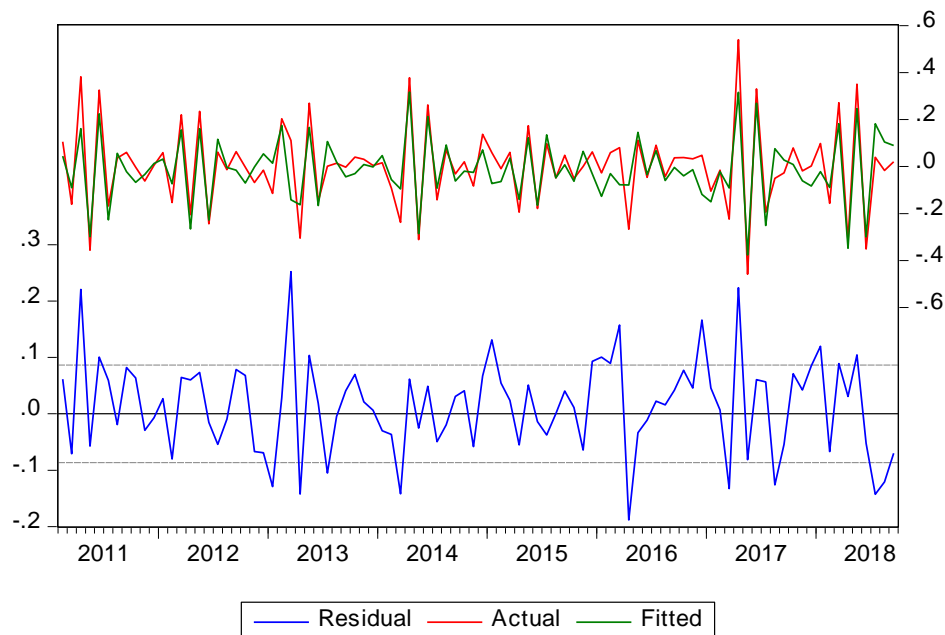
Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.*.	.*.	1	-0.110	-0.110	1.1520	
. .	. .	2	0.011	-0.002	1.1628	
.*.	.*.	3	-0.068	-0.068	1.6141	
.*.	.*.	4	-0.163	-0.181	4.2267	0.040
. .	. .	5	0.023	-0.018	4.2788	0.118
. *	. *	6	0.077	0.076	4.8790	0.181
. .	.*.	7	-0.059	-0.071	5.2349	0.264
.*.	.*.	8	-0.120	-0.175	6.7192	0.242
. .	. .	9	0.010	-0.011	6.7298	0.347
.*.	.*.	10	-0.107	-0.096	7.9410	0.338
. *	. *	11	0.193	0.128	11.901	0.156
. .	. .	12	-0.022	-0.044	11.953	0.216
. *	. *	13	0.168	0.171	15.044	0.130
. .	. .	14	-0.046	-0.005	15.280	0.170
. .	. .	15	-0.018	0.017	15.317	0.225
. .	. *	16	0.063	0.079	15.763	0.262
. .	. .	17	-0.005	0.043	15.765	0.328
.*.	.*.	18	-0.070	-0.088	16.330	0.360
.*.	.*.	19	-0.077	-0.075	17.031	0.384
. .	. .	20	-0.039	-0.022	17.209	0.440
. .	. .	21	-0.057	-0.003	17.605	0.482
. .	.*.	22	-0.006	-0.122	17.609	0.549
. *	. *	23	0.096	0.113	18.774	0.537
.*.	.*.	24	-0.143	-0.201	21.367	0.437
. .	. .	25	0.033	0.004	21.508	0.490
. .	. .	26	0.012	-0.059	21.528	0.549
. *	. *	27	0.133	0.166	23.871	0.469
. .	. .	28	0.072	0.007	24.577	0.486
.*.	.*.	29	-0.087	-0.119	25.614	0.484
.*.	.*.	30	-0.110	-0.096	27.297	0.448
.*.	. .	31	-0.086	-0.005	28.349	0.446
.*.	.*.	32	-0.099	-0.190	29.775	0.425
. .	. .	33	0.003	-0.037	29.776	0.477
. *	. .	34	0.088	-0.014	30.923	0.470
.*.	. .	35	-0.148	-0.036	34.255	0.360
. **	. *	36	0.231	0.128	42.461	0.125

Source: output of statistical software EViews 9

Figure 5. shows actual vs. fitted values and residuals. As we can see from the figure the model fits quite well to the actual values. The pattern of the residuals seems to confirm the model to be adequate.

Figure 5: **Monthly tourist arrivals to Croatia – actual, fitted and residual**



Source: output of statistical software EViews 9

## CONCLUSION

Tourism is an important sector of Croatian economy, generating fifth of its GDP. For the purpose of tourism development, it is very important to have quality research in the field as an input for strategic planning. One of the most fundamental need for further development of tourism is demand forecasting. This paper aimed to give its contribution to answering that specific need. Tourism demand forecasting methods can be divided in qualitative and quantitative methods. Quantitative methods used for forecasting purposes are either time series models or econometric studies. In this paper tourism demand for Croatia has been modelled using Box and Jenkins seasonal ARIMA methodology. The Box-Jenkins methodology refers to a set of procedures for identifying and estimating time series models within the class of autoregressive integrated moving average (ARIMA) models. The data for tourist arrivals showed clear patterns of seasonality and an upward trend. After differentiation and seasonal adjustment, a model selection and diagnostic checking followed. The model selected was seasonal ARIMA (2,1,1) (0,1,1)<sub>12</sub>. The model fits the observed data well and shows no autocorrelation of the residuals. The value of the model developed in this paper is in its practical implications for tourism demand forecasting. The information provided from econometric models can

be of great value for decision makers such as local government, tourist boards, tourism ministry and tourism companies. Future research should aim in developing models using other methodologies mentioned in this paper. Comparative analysis of their performance is also recommended.

## REFERENCES

- Baldigara, T. and Koić, M., (2015), "Modelling the international tourism demand in Croatia using a polynomial regression analysis", *Turističko poslovanje*, No. 15, pp. 29-38.
- Biljana, P., (2017), "Forecasting international tourism demand: the evidence from Macedonia", *UTMS Journal of economics*, Vol. 3, No. 1, pp. 45-55.
- Box, G.E.P and Jenkins, G.M. (1976), *Time series analysis: forecasting and control*, San Francisco, Holden – Day.
- Cerović, Z., Grudić Kvasić, S. and Ivančić, I. (2017), *Forecasting tourism demand – the case of the city of Rijeka*, DIEM, Vol. 3, No. 1, pp. 750-763.
- Chhorn, T. and Chaiboonsri, C. (2018), "Modelling and Forecasting Tourist Arrivals to Cambodia: An Application of ARIMA – GARCH Approach", *Journal of Management, Economics, and Industrial Organization*, Vol. 2, No. 2, pp.1-19.
- Constantino et al. (2016), "Tourism demand modelling and forecasting with artificial neural network models: The Mozambique case study", *Tekhne*, Vol. 14, No. 2, pp. 113-124, <https://doi.org/10.1016/j.tekhne.2016.04.006>
- Han, Z., Durbarry, R., and Sinclair, M. T. (2006), "Modelling US tourism demand for European Hong Kong tourism", *International Journal of Hospitality Management*, Vol. 22, pp. 435-451.
- Hyndman, R.J. and Kostenko, A.V. (2007), "Minimum sample size requirements for seasonal forecasting models", *Forsight*, 6, pp. 12-15.
- Lee, K.N. (2010), "Forecasting long – haul tourism demand using error correction model", *Applied Economics*, Vol. 43, No. 5, pp. 527-549.
- Sakhuja, S., Jain, V., Kumar, S., Chandra, C., Ghildayal, S.K. (2016), "Genetic algorithm based fuzzy time series tourism demand forecast model", *Industrial Management & Data Systems*, Vol. 116, Issue 3, pp. 483-507, <https://doi.org/10.1108/IMDS-05-2015-0165>
- Singh, E.H. (2013), "Forecasting tourist inflow in Bhutan using seasonal ARIMA", *IJSR*, Vol. 9, No. 2, pp. 242-245.
- Song, H. and Li, G. (2008), "Tourism demand modelling and forecasting – A review of recent research", *Tourism Research*, Vol. 29, No. 2, pp. 203-220,
- Song, H. and Witt, S.F. (2006), "Tourism demand forecasting: a time varying parameter error correction model", *Journal of Travel Research*, Vol. 45, No. 2, pp. 175-185.
- Song, H., Li, G., Witt, S.F. and Athanasopoulos, G. (2011), "Forecasting tourist arrivals using time varying parameter structural time series models", *International Journal of Forecasting*, Vol. 27, No. 3, pp. 855-869.
- Song, H., Wong, K.K.F., and Chon, K.K.S. (2003c), "Modelling and forecasting the demand for destinations", *Tourism Management*, 27, pp. 1-10.
- Tica, J. and Kožić, I. (2015), "Forecasting Croatian inbound tourism demand", *Economic Research*, Vol. 28, No. 1, pp. 1046-1062.
- Turner, L.W., and Witt, S.F. (2001a), "Factors influencing demand for international tourism: Tourism demand analysis using structural equation modelling, revisited", *Tourism Economics*, 7, pp. 21-38.
- [www.dzs.hr](http://www.dzs.hr)

**Damira Đukec**, PhD, Assistant Professor  
University North  
Department of Business Economy  
104. brigade 3, 40 000 Varaždin, Croatia  
Phone: 091 159 3575  
E-mail: [ddukec@unin.hr](mailto:ddukec@unin.hr)