CIVITAS, 2022, 12(1), 26-38
www.civitas.rs

Cvetko Andreeski[1]
UDC 338.48: :[616.98:578.834(100)"2020/2021"
Original research article
Received: 02/01/2022
Accepted: 11/03/2022

TOURISM TIME SERIES ANALYSIS CONTAINING STRUCTURAL CHANGES DUE TO THE COVID-19 PANDEMIC: SELECTING THE OPTIMAL MODEL

Abstract: Tourism is exposed to various risks such as natural disasters, different types of crises, negative propaganda, pandemics, and the like. All these risks affect not only the development of tourism, but also time series models that are subject to analysis. In order to be able to make informed decisions that influence the functioning and further development of the tourism sector, it is necessary to make an accurate analysis of the current situation and to be able to predict the future values of the time series model. Due to the risks stated above, this complex problem requires an in-depth analysis, a selection of an appropriate model and testing the model's results and predictions. The paper examines time series analyses in the field of tourism and structural changes before and during the pandemic, as well as possible models that can be used to model such series in the future.

Keywords: time series, structural changes, analysis, models, tourism

1. Introduction

Tourism has been one of the most affected economic activities during the COVID-19 pandemic. With the outbreak of the pandemic in Europe in March 2020, the catering and accommodation facilities were

[1] Full Professor, Faculty of Tourism and Hospitality Ohrid, email: cvetko.andreeski@uklo.edu.mk

closed, preventing the regular functioning in this sector. According to the World Bank[2], in 2019, the year before the pandemic outbreak, the export effect of tourism in Serbia amounted to 7.65% of GDP. In 2020, however, the total number of overnight stays was only 24% of the overnight stays in 2019, and the revenues dropped by more than 147 million dollars.

Significant and long-lasting structural changes in tourism development during the pandemic has pushed the sector to the brink of sustainability; furthermore, this prevented researchers from modelling time series analyses for tourism development and predicting future values (O'Hare & Li, 2015; Asghar & Amena, 2012). There have been many studies aimed at developing a valid model for identifying and predicting future time series data in the field of tourism development (Andreeski & Mechkaroska, 2020; Petrevska, 2017; Baldigara & Mamula, 2015), but these models seem useless in the face of a momentous structural change such as the COVID-19 pandemic. Besides linear models for identification and prediction, more complex non-linear models are used, such as the models based on artificial neural networks, which can include the changes in the level and variance of the series (Andreeski & Petrevska, 2021; Shi, 2019). According to Boot & Pick (2019), forecasting can be slightly improved if a post-break sample rather than the full sample is used. However, this paper will consider only small breaks, not lasting ones. Hoping to get more accurate results, some researchers have used the deep-learning and big-data analysis techniques; still, all these models are structured on the assumption that future values will be based on historical trends and variance, i.e., that future trends will reflect historical ones (Kaushnik et al., 2019; Jian et al., 2021). However, even these models cannot provide valid results if a long enough data series from the beginning of the pandemic is unavailable. Some studies of modelling and predicting tourism time series during the pandemic examine the current situation and offer comparison with predicted values based on the pre-pandemic tourism development series, with the aim to calculate the losses incurred by the pandemic in this sector (Andreeski & Petrevska, 2021; Šenkova et al., 2021). Recently, some publications have proposed models for identifying and predicting the future values of tourism development series (Provenzano & Volo, 2021; Šenkova et al., 2021).

[2] *https://data.worldbank.org/indicator/ST.INT.TRNX.CD?locations=RS*

Nearly two years since the outbreak of the pandemic, enough time has passed for a time series analysis modelling to be attempted, so that future data may be predicted. According to the recommendations for modelling a series of sufficient data, the time required is at least four years, provided monthly data is available. In this paper one such model is proposed, based on the linear SARIMA model, during which some pre-processing of the time series is done. The model was chosen because of its relative simplicity and ability to compare this model's performance with that of similar models used in studies on modelling time series in the field of tourism.

The model was created using the data on overnight stays of foreign tourists in the Republic of Serbia ten years before the pandemic and one year after the pandemic.

2. Data and Modelling

Fig. 1 shows the number of overnight stays of foreign tourists in the Republic of Serbia 2010-2019, as well as the numbers for January - September 2021. Data presented in Fig. 1 are taken from the Statistical Office of the Republic of Serbia[3].

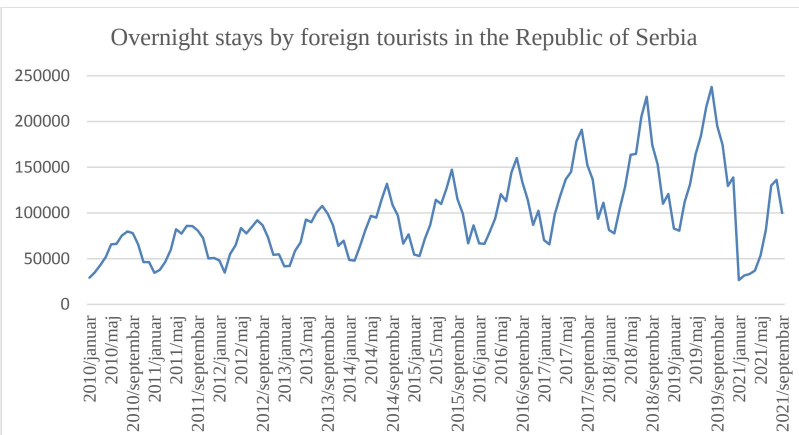


Fig 1. Number of overnight stays by foreign tourists in the Republic of Serbia

[3] *https://data.stat.gov.rs/Home/Result/220203?languageCode=en-US*

From the displayed data, several characteristics of the series are evident:

- The series has a defined trend in the 2010-2019 period, then shows a significant decline in 2021, but the shape of the series remains constant.
- The series has a change in variance, i.e., a pronounced heteroskedasticity.
- The seasonal characteristic of the series is obvious in the entire analysed period.
- Probable structural change in 2021, due to the trend change.

The series deliberately omits the 2020 data, because that year should not be taken as a reference for the number of tourist arrivals, as it would make modelling the series and predicting the future data impossible.

The SARIMA model was used for modelling the series. In the pre-processing part of the series, a logarithm of the original series was made, and after that a differentiation of the series in order to get a stationary series that can be modelled. A unit root test was performed for this modified series in order to identify stationarity. Table 1 shows the result of this test.

Table 1. Unit root test of the modified series

Null Hypothesis: NOCENJA_STRANI_DLOG has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, max. lag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-19.04046	0.0000
Test critical values:	1% level	-3.491928	
	5% level	-2.888411	
	10% level	-2.581176	

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

Table 1 shows that the value of the Augmented Dickey Fuller test is significantly lower than the critical values for the stationary series and the probability of rejecting the hypothesis that the series is stationary is less than 1%.

Before moving on to modelling the series, it is necessary to identify the structural change that is likely to occur in 2021. For that purpose, a unit root break test of the series was made. The test results are given in Table 2.

Table 2. Unit root break test of the modified series

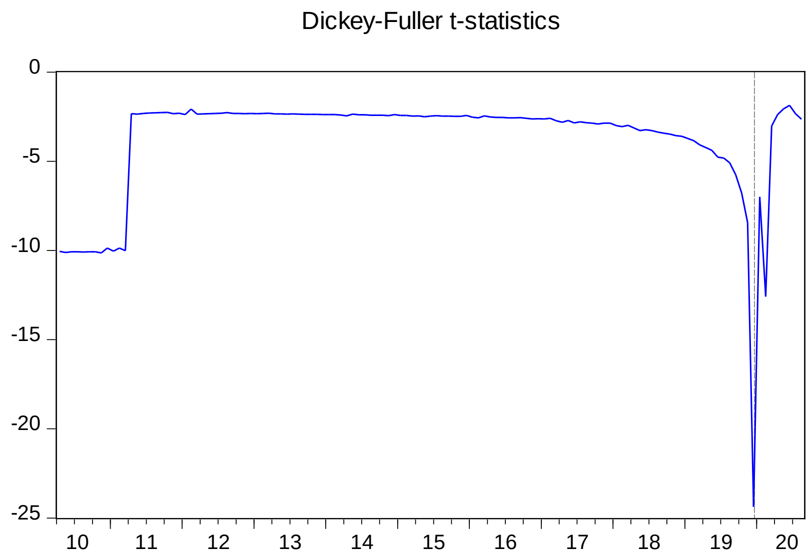
Null Hypothesis: NOCENJA_STRANI_DLOG has a unit root
Trend Specification: Intercept only
Break Specification: Intercept only
Break Type: Innovation outlier

Break Date: 2019M12
Break Selection: Minimize Dickey-Fuller t-statistic
Lag Length: 10 (Automatic - based on Schwarz information criterion, Max. lag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-24.36267	< 0.01
Test critical values:	1% level	-4.949133	
	5% level	-4.443649	
	10% level	-4.193627	

From the results given in Table 2, it can be concluded that there is a structural change of the series at the end of 2019, i.e., the structural change of the series after 2019 is detected.

Fig. 3. Graphical representation of structural change according to Dickey Fuller t-statistics



After stationing the series and testing the existence of structural changes, modelling can be resumed. To determine the independent variables of the series, a correlation of the series was made. The values of the correlogram are given in Table 3.

Table 3. Correlogram of the modified series

Sample: 2010M01 2020M09
Included observations: 128

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.100	0.100	1.3171	0.251
		2	0.191	0.183	6.1325	0.047
		3	-0.003	-0.039	6.1336	0.105
		4	-0.160	-0.200	9.5639	0.048
		5	-0.229	-0.209	16.671	0.005
		6	-0.585	-0.553	63.330	0.000
		7	-0.180	-0.165	67.799	0.000
		8	-0.127	0.027	70.022	0.000
		9	-0.066	-0.103	70.630	0.000
		10	0.070	-0.168	71.316	0.000
		11	0.137	-0.201	74.001	0.000
		12	0.603	0.365	126.16	0.000

The correlogram shows that the sixth and 12th delays have the highest values, which indicate the seasonal component of the series. It is necessary to test the serial correlation in the series, but this is unlikely because the first delay has a small value.

Several different competing models have been made for the series, and only the ones that give the best results are shown in the paper. The first model has only one independent variable, the 12th delay in the series. The results are shown in Table 4.

Table 4. Model of the series for tourism overnight stays in the Republic of Serbia

Dependent Variable: NOCENJA_STRANI_DLOG
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 11/07/21 Time: 09:21
Sample: 2010M02 2020M09
Included observations: 128
Convergence achieved after 11 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR (12)	0.877828	0.072841	12.05135	0.0000
SIGMASQ	0.023934	0.001189	20.12629	0.0000
R-squared	0.642378	Mean dependent var		0.009632
Adjusted R-squared	0.639539	S.D. dependent var		0.259714
S.E. of regression	0.155928	Akaike info criterion		-0.725325
Sum squared resid.	3.063501	Schwarz criterion		-0.680762
Log likelihood	48.42082	Hannan-Quinn criter.		-0.707219
Durbin-Watson stat	2.283420			

The results show that the independent variable is valid, i.e., it has a high value of t-statistics and a low probability that this parameter will be removed from the model. The degree of variance of the original series is about 64%, and the value of the Durbin-Watson statistic is close to 2, indicating that there is no significant serial correlation of the residuals.

In order to improve the model, a dummy variable is made that aims to model the structural change. This variable has a value of 1 for the 2021 months, and for data in previous years it has a value of 0. The modelling results of this model are shown in Table 5.

From the results given in Table 5 the following can be concluded:

- The Dummy variable is relevant to the model and the probability of ejecting this parameter from the model is less than 1%.
- This model covers most of the model variance.
- The information criterion has a lower absolute value compared to the previous model.

The value of Durbin-Watson statistics has a value closer to 2 compared to the previous model. A correlation of residuals was made for the model which shows that all residuals are within the confidence interval ± 2se. The forecast of the future values of the time series is given below, with a forecast of the future values based on the selected model.

Table 5. Model of the series for tourism overnight stays in the Republic of Serbia with added dummy variable

Dependent Variable: NOCENJA_STRANI_DLOG

Method: ARMA Maximum Likelihood (OPG - BHHH)

Date: 11/07/21 Time: 09:26

Sample: 2010M02 2020M09

Included observations: 128

Convergence achieved after 7 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DUMMY	-0.083368	0.026179	-3.184512	0.0018
AR(12)	0.882392	0.071973	12.26012	0.0000
SIGMASQ	0.023367	0.001553	15.04329	0.0000
R-squared	0.650839	Mean dependent var		0.009632
Adjusted R-squared	0.645253	S.D. dependent var		0.259714
S.E. of regression	0.154687	Akaike info criterion		-0.730304
Sum squared resid	2.991016	Schwarz criterion		-0.663460
Log likelihood	49.73946	Hannan-Quinn criter.		-0.703145
Durbin-Watson stat	2.264859			

3. Forecasting of the Series

To predict the series, an extension of the series was made for 12 months and the return of the predicted values to correspond to the values of the original series. For research purposes, an in-sample forecast of the values has been made, for the known values from 2021, as shown in Fig. 4.

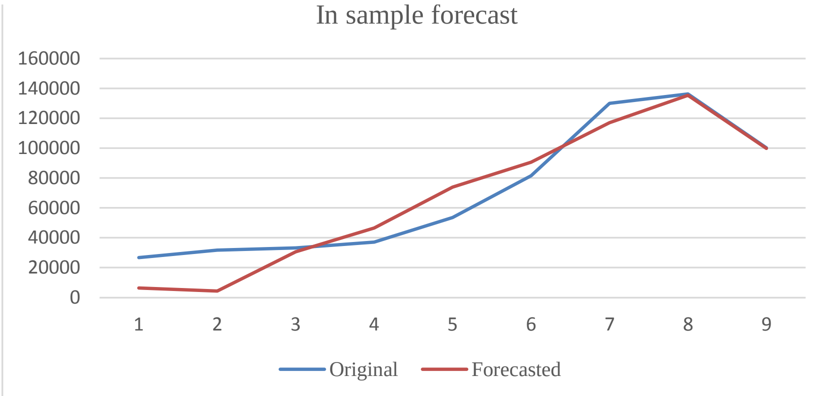


Fig 4. In sample forecast of the values for 01/09/2021

Fig 4. In sample forecast of the values for 01/09/2021

Finally, the values for the expected number of foreign tourists for 2022 are forecasted, again for the first nine months. These results are shown in Fig 5. The same graph shows the series of values for 2021, to compare the two series.

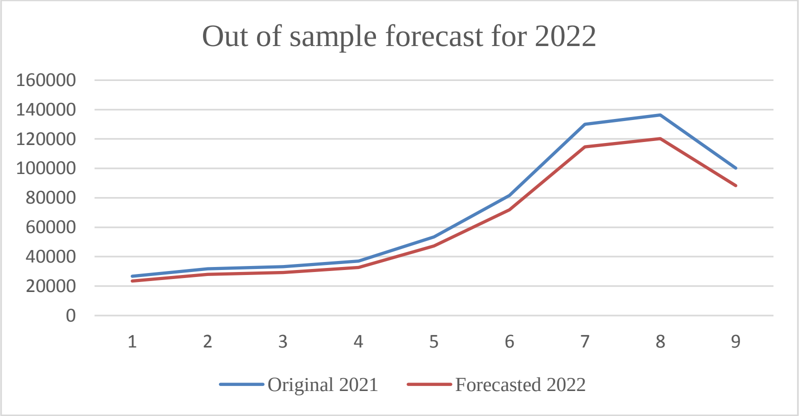


Fig 5. Out of sample forecast of the values for 01/09/2022

The dummy variable was not used to calculate the values in 2022, because another structural change for this year is not likely to occur. Otherwise, no prediction could have been made at at all. As the graph shows, a smaller number of tourists is forecasted for 2022 compared to 2021, which is highly unlikely. However, the model does the calculation based on the entire history for the analysed series and the value of the ar (12) parameter is less than 1, which means that future values should have lower values than the previous ones. The estimated decline in overnight stays is around 8%.

4. Conclusions

The analysis uses a time series of data on overnight stays by foreign tourists in the Republic of Serbia. This is a relevant time series model, which can be used to monitor the COVID-19 pandemic on tourism development in any country. The series has a structural change that makes modelling the series more complex. By properly pre-processing the series and adding a dummy variable to the modelling, a valid model has been created. Finally, the forecast of the future values of the series shows lower values for the next year (2022) compared to those of the previous year (2021), which is likely to occur. However, the importance of this model is that it has shown that it is possible to make a valid model even for series which are a challenge for modelling. When there are enough data from the series after the structural change, it will be possible to create a more reliable model based on the post-pandemic data. This model is expected to provide more accurate predictions of future values.

References

Andreeski, C. & Mechkaroska, D., (2020). Modelling, Forecasting and Testing Decisions for Seasonal Time Series in Tourism. *Acta Polytechnica Hungarica 17 (10),* pp. 149-171.

Andreeski, C. & Petrevska, B., (2021). *FORECASTING DYNAMIC TOURISM DEMAND USING ARTIFICIAL NEURAL NETWORKS.* Skopje, s.n.

Asghar, Y. & Amena , U., (2012). Structural Breaks, Automatic Model Selection and Forecasting Wheat and Rice Prices for Pakistan. *Pakistan Journal of Statistics and Operation Research,* pp. 1-20.

Baldigara, T. & Mamula, M., (2015). MODELLING INTERNATIONAL TOURISM DEMAND. *Tourism and Hospitality Management, Vol. 21, No. 1.,* pp. 19-31.

Boot, Tom; Pick, Andreas (2019). Does modeling a structural break improve forecast accuracy?. *Journal of Econometrics, (), S0304407619301824–*. doi:10.1016/j.jeconom.2019.07.007

https://data.stat.gov.rs/Home/Result/220203?languageCode=en-US. [Online]

https://data.worldbank.org/indicator/ST.INT.TRNX.CD?locations=RS. [Online]

Jian, B. W., Hui, L. & Zhi, F. P., (2021). Tourism demand forecasting with time series imaging: A deep learning model. *Annals of Tourism Research,* Volume 90.

Kaushnik, R., Jain, S. & Dash, T., (2019). Performance evaluation of deep neural networks for forecasting time-series with multiple structural breaks and high volatility. *Computer Science, Mathematics (arXiv preprint arXiv:1911.06704).*

O'Hare, C. & Li, Y., (2015). Identifying Structural Breaks in Stochastic Mortality Models. *SSRN electronic Journal* .

Petrevska, B., (2017). Predicting tourism demand by ARIMA models. *Economic research-Ekonomska istraživanja,* pp. 939-950.

Provenzano, D. & Volo, S., (2021). Tourism recovery amid COVID-19: The case of Lombardy, Italy. *Tourism Economics, https://doi.org/10.1177/13548166211039702.*

Šenkova, A. et al., (2021). Time Series Modeling Analysis of the Development and Impact of the COVID-19 Pandemic on Spa Tourism in Slovakia. *Sustainability 13(20) 11476.*

Shi, X., (2019). Tourism culture and demand forecasting based on BP neural network. *Personal and Ubiqutous Computing,* pp. 299-308.