DOI 10.20544/HORIZONS.A.23.2.18.P01
UDC: 338.482:519.246.8(497.7)"2010/2017"

CHALLENGES ON MODELING TIME SERIES OF BASIC TOURIST PARAMETERS IN THE REPUBLIC OF MACEDONIA[1]

Cvetko Andreeski
University "St. Kliment Ohridski" – Bitola, Faculty of Tourism and Hospitality,
cvetko.andreeski@yahoo.com
Marina Blazekovic Toshevski
University "St. Kliment Ohridski" – Bitola, Faculty of Information and Communication Technologies, marina.blazekovic@yahoo.com

ABSTRACT

The basic parameters in development of tourism in the Republic of Macedonia, number of arrived tourists and number of spent nights have accentuated seasonal component and trend for foreign tourist. Besides the fact that these series have seasonal components, in some cases we can find structural break in the series. All these obstacles can be overcome if we choose appropriate model for basic tourist parameters. In these paper we present some of the series from regions in the Republic of Macedonia. We have chosen relevant regions that affects the development of tourism in the Republic of Macedonia. We made models for the series and also we have tested all the models for validity and stability. For created model we have tested the model with in sample and out of sample forecast of the series.

KEY WORDS: time series, analysis, tourism, forecast.

[1] Original scientific paper

INTRODUCTION

Tourism time series gives us important information about the trends in tourism, seasonality issues in the series, but also the changes in the attitudes of the guest, as well as the attractiveness of the destination. As the tourism has very dynamic behavior in some period of development of destination, sometimes it's not that easy to create appropriate model for tourism time series. On the other side, we need to have valid models and forecast for the new season in order to make good decisions and to get ready and prepare for new arrivals and new number of basic tourist parameters.

As the time series analysis is very important for tourism, we can find many papers that are dealing with modeling of these types of time series. In the paper (Petrevska, Biljana;, 2015) modeling and forecasting is made on basic tourism parameters for the period 1953-2014 for yearly data. This series is real challenge for modeling because it has two detected breaks (structural changes). For the modelling she uses standard ARIMA model as well as we can find at (Baldigara & Mamula, 2015). In the paper (Andreeski & Vasant, 2009) we can find the comparison between linear ARIMA models and non-linear models based on artificial neural networks. In the paper we can find exploration of their performances for modeling on time series with included break(s). Many authors have worked on time series modeling on time series with seasonal components by using neural networks мрежи(Benkachca , Behra, & El Hassani, August 2013) (Chen, 2011) (Oscar, Enric , & Salvador, 2013). Time series with significant seasonal component are important in different areas of research (Shengwei, Juan, & Gang, 2013), (Ette, 2012).

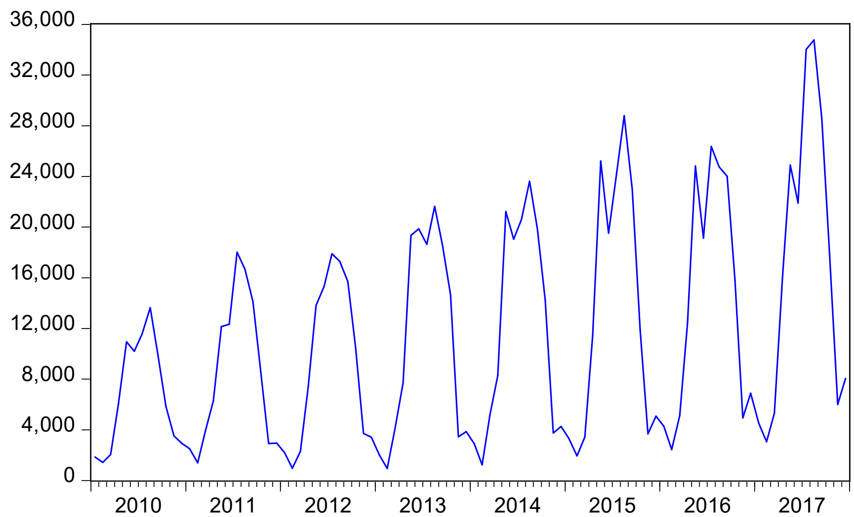
In this paper we have two time series for arrived tourist in South-West planning region in the Republic of Macedonia. This region participates in the total number of spend nights in the Republic of Macedonia on yearly basis for more than 50%, so it's very important for the development of the tourism. In last ten years we can see that we have increasing trend in the development of tourism, especially for the number of arrived foreign tourists.

In further text we have modeling of the two time series, analysis on the behavior of the series, and forecasting of future values. At the end we have made conclusions based on the results of analysis.

MODELING A SERIES OF FOREIGN TOURISTS IN THE SOUTH-WEST PLANNING REGION (2010-2017)

Model series is the series of arrived foreign tourists in the South-west planning region. This region, according to the number of arrived tourists, is second, but very important given that in this region are the most visited tourist places during the summer season. The Ohrid coast in the summer period participates with more than 50% of the total number of arrived tourists in the Republic of Macedonia. Chart 1 shows the series of arrived foreign tourists in the South-west planning region in the period January 2010 - December 2017.

Chart 1. Graphic presentation of incoming foreign tourists January 2010 - December 2017



From the shown series, it can be noticed that the series has a seasonal component, there is a pronounced upward trend in the considered period, but there is also a change in the behavior of the series, that is, expressed heteroscedasticity. As of 2013, there are fewer tourists in June compared to May. Has been made an analysis of the stability of the series, i.e. checking whether the series has a structural change and according to the obtained results, it has been demonstrated that there has been a structural change in the series in 2013. The results of the analysis of structural change are given in Table 1.

Table 1. Presentation of the results of the analysis of structural changes in the series

Null Hypothesis: STRANSKI_JZPR_LOG has a unit root
Trend Specification: Intercept only
Break Specification: Intercept only
Break Type: Innovational outlier

Break Date: 2013M02
Break Selection: Minimize Dickey-Fuller t-statistic
Lag Length: 11 (Automatic - based on Schwarz information criterion, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.506874	> 0.99
Test critical values:	1% level	-4.949133	
	5% level	-4.443649	
	10% level	-4.193627	

Based on the results of the analysis of structural change, it is necessary to make a model for the series either after structural change, or by a method that can do modeling of the series with built-in structural change. Given that from the beginning of 2013, until December 2017 has more than 50 values of the series, we made an attempt to model this period of the series. The further results of the analysis were made for the series from January 2013 December 2017.
Table 2 shows the results of the single roots test for the original series of foreign tourists 2010-2017.

Table 2.Test of single roots for checking stationarity of the original series of foreign tourists in SWPR

Null Hypothesis: STRANSKI_JZPR has a unit root
Exogenous: Constant
Lag Length: 11 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.200432	0.9711
Test critical values:	1% level	-3.510259	
	5% level	-2.896346	
	10% level	-2.585396	

From the results in Table 2 it can be concluded that the probability of accepting the null hypothesis for the single root of the series is very high, which means that the analyzed series is non-stationary and it is necessary to differentiate the original series.
Table 3 gives the results of the single roots test for a differentiated series of foreign tourists in SWPR

Table 3.Test of single roots for checking the stationarity of the differentiated series of foreign tourists in SWPR

Null Hypothesis: D(STRANSKI_JZPR) has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-10.70429	0.0001
Test critical values:	1% level	-3.510259	
	5% level	-2.896346	
	10% level	-2.585396	

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

According to the result of single roots presented in Table 24, the likelihood of accepting the null hypothesis for the existence of a single root of the series is zero, that is, the differentiated series is stationary and can be used to create a model. However, due to the heteroscedasticity of the model, additional modification of the original series was made, i.e. the logarithm of the series was made in order to obtain a time series with a smaller variance change over time. The single roots test for such a modified series and the first series differentiation is given in Table 4.

Table 4.Test of single roots for checking stationarity of logarithmic and differentiated series of foreign tourists in SWPR

Null Hypothesis: D(STRANSKI_JZPR_LOG) has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-19.46932	0.0001
Test critical values:	1% level	-3.514426	
	5% level	-2.898145	
	10% level	-2.586351	

The resulting series has a much higher value for t statistics, but a smaller change in the variance. For this reason, we will use this modified series to create a model of the series.

For the modeling of the series was made a correlogram of delays. The results of the correlogram of the differentiated series, are given in Table 5.

From the correlation analysis we can conclude the following:

- The model has an emphasized seasonal component annually
- The model has an emphasized seasonal component at the semi-annual level
- The model has an emphasized serial correlation of delays

Table 5. Correlation analysis of the bites of the series SWPR - foreign tourists 2013-2017

Date: 04/26/18 Time: 22:14
Sample: 2013M03 2017M12
Included observations: 54

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.228	0.228	2.9698	0.085
		2	0.083	0.033	3.3698	0.185
		3	0.058	0.034	3.5679	0.312
		4	-0.116	-0.146	4.3783	0.357
		5	-0.255	-0.219	8.3953	0.136
		6	-0.598	-0.556	30.910	0.000
		7	-0.205	-0.019	33.605	0.000
		8	-0.231	-0.245	37.112	0.000
		9	0.017	0.184	37.131	0.000
		10	-0.009	-0.282	37.136	0.000
		11	0.178	0.095	39.375	0.000
		12	0.718	0.488	76.478	0.000

Accordingly, the model should include the three segments to create a good model for the given time series. As a well-balanced model is proposed a model with an autoregressive independent variable ar(12) and an independent variable with moving averages ma(1). Components with a half-year delay did not yield good results when included in the model. The modeling results are given in Table 6.

Table 6. Results of the modeling of the series of foreign tourists for SWPR

Dependent Variable: STRANSKI_JZPR_LOGD
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/26/18 Time: 22:15
Sample: 2013M03 2017M08
Included observations: 54
Convergence achieved after 15 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	0.985614	0.011843	83.22057	0.0000
MA(1)	-0.964937	0.038963	-24.76517	0.0000
SIGMASQ	0.005014	0.000962	5.215059	0.0000

R-squared	0.935948	Mean dependent var	0.028997
Adjusted R-squared	0.933437	S.D. dependent var	0.282422
S.E. of regression	0.072865	Akaike info criterion	-1.545645
Sum squared resid	0.270771	Schwarz criterion	-1.435146
Log likelihood	44.73241	Hannan-Quinn criter.	-1.503030
Durbin-Watson stat	1.761174		

From the results of the model we can conclude the following:

- The probability of ejection of any of the independent variables is small
- The extent to which the model follows the original series is high, which promises a good forecast of the future values of the series
- The value of Durbin-Watson's statistics is close to 2, indicating that there is little likelihood of the serial correlation of the residuals
- Information criteria indicate that there may be a better model, but given that we do not have a competitive model for this series, we will not consider these criteria

Table 7 gives the values of the correlation of the residuals.

Table 7.Correlogram of the residuals for the model in Table 6.

Date: 04/26/18 Time: 22:59
Sample: 2013M03 2017M12
Included observations: 54
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.050	0.050	0.1422	
		2	0.065	0.063	0.3878	
		3	-0.065	-0.072	0.6410	0.423
		4	-0.099	-0.097	1.2286	0.541
		5	-0.112	-0.096	2.0003	0.572
		6	-0.110	-0.095	2.7607	0.599
		7	0.108	0.121	3.5124	0.622
		8	-0.014	-0.032	3.5247	0.741
		9	0.149	0.110	5.0082	0.659
		10	0.050	0.028	5.1787	0.738
		11	0.093	0.075	5.7874	0.761
		12	-0.210	-0.213	8.9626	0.536

From the results in Table 7 we can conclude that the residuals are an uncorrelated, that the values of the correlation are within the standard interval of confidence and that this series can be defined as a series of white noise.

In addition to having made the change of the series with logarithmic function and one differentiation, model has two independent variables.The original series is non-stationary and it has an upward trend, a built-in structural change that prevents modeling of the series throughout the analyzed period 2010-2017.

At the end of the analysis of this series, we made forecast of the future values of the series within the series itself, but also for the period of one year after the last value in the series, December 2017.

Figure 1 presents the model and the original series for the period July 2017 - December 2017.

The result of the forecast is somewhat weaker than was the case with the previous models. The reason for this is the complexity of the model, but also the amount of data on which the model is made. We expect, in case there will be no other structural changes in the future, the model itself will show better results. For this series it is difficult to make another model of identification and forecast.We can use non-linear models which can identify series with included structural changes, but they require a rich input of data.

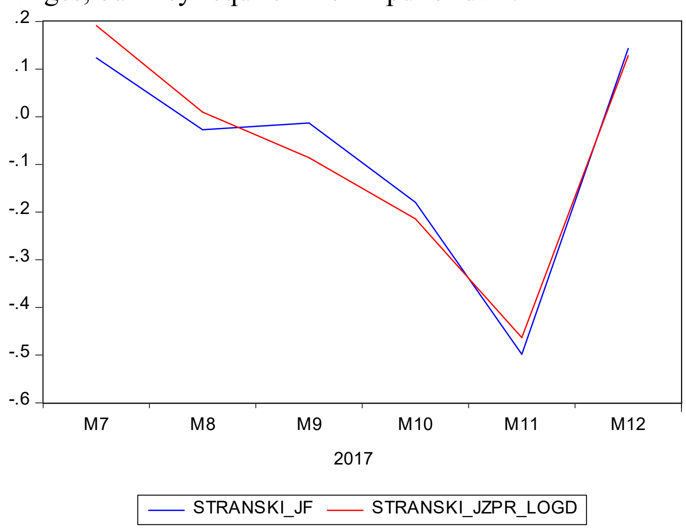


Figure 1. Model presentation within the February-July 2017 series

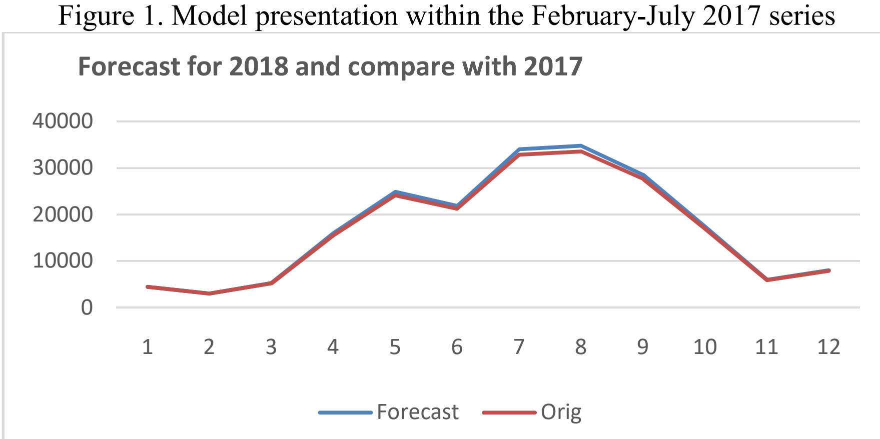


Figure 2. Forecast of the model beyond the scope of the January 2018 - December 2018 series

MODELING A SERIES OF DOMESTIC TOURISTS IN THE SOUTH-WEST PLANNING REGION (2010-2017)

For the Southwest planning region we've made analysis and modeling of the time series of the number of domestic tourists for the same period 2010-2017. The graphical display of the series is given in Chart 2.

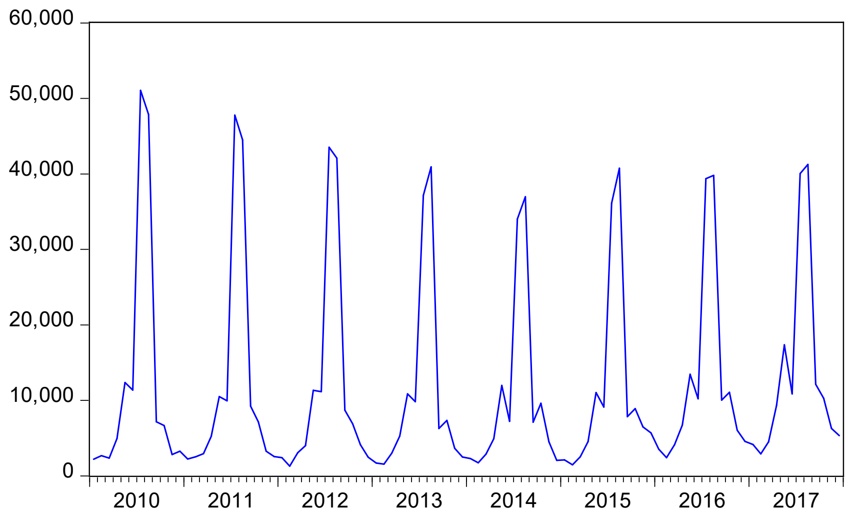


Chart 2. Graphic presentation of the time series of incoming domestic tourists, 2010-2017

From the graphic it can be noted that:

- The series has no defined trend over the entire period. By 2014 there is a downward trend in the number of arrived tourists, and then comes a period of stabilization of the number
- It can not be said that there is a tendency of changing the variance throughout the analyzed period. We will try to create a model with minimal transformation of the original series in order to obtain a stationary time series
- The series has an emphasized seasonal component, much more pronounced than is the case with the time series of arrived foreign tourists in the same period

Table 8 gives the test of single roots for the different series of incoming domestic tourists.

Table 8. Single roots test for a differentiated series of incoming domestic tourists

Null Hypothesis: D(DOMASNI_JZPR) has a unit root
Exogenous: Constant
Lag Length: 11 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.622832	0.0000
Test critical values:	1% level	-3.511262	
	5% level	-2.896779	
	10% level	-2.585626	

From the results it can be concluded that the differentiated series is stationary and that one differentiation is sufficient for stationing the series.
To create a model of the analyzed series was created a correlogram. The results of the correlogram are given in Table 9.

Table 9.Correlogram of the differentiated series of domestic tourists

Date: 04/29/18 Time: 18:19
Sample: 2010M01 2017M12
Included observations: 95

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.031	-0.031	0.0940	0.759
		2	-0.323	-0.325	10.446	0.005
		3	0.078	0.061	11.054	0.011
		4	-0.136	-0.266	12.925	0.012
		5	-0.050	-0.011	13.183	0.022
		6	-0.069	-0.260	13.678	0.033
		7	-0.047	-0.060	13.910	0.053
		8	-0.115	-0.364	15.312	0.053
		9	0.075	0.023	15.917	0.069
		10	-0.275	-0.813	24.088	0.007
		11	-0.036	-0.133	24.232	0.012
		12	0.844	0.313	103.33	0.000

From the correlogram it can be concluded that the series has a pronounced seasonality, but also a serial partial correlation. Accordingly, has been created a proposed model that covers the characteristics of the series. The results of the model are given in Table 10.

Table 10. Model of the series for incoming domestic tourists

Dependent Variable: DOMASNI_JZPR_D
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/29/18 Time: 18:22
Sample: 2010M02 2017M12
Included observations: 95
Convergence achieved after 41 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	0.994059	0.001560	637.2150	0.0000
MA(1)	-0.891582	0.048254	-18.47675	0.0000
SIGMASQ	2812649.	353389.8	7.959056	0.0000

R-squared	0.984856	Mean dependent var	33.05263
Adjusted R-squared	0.984527	S.D. dependent var	13700.59
S.E. of regression	1704.220	Akaike info criterion	18.32157
Sum squared resid	2.67E+08	Schwarz criterion	18.40222
Log likelihood	-867.2747	Hannan-Quinn criter.	18.35416
Durbin-Watson stat	1.448561		

From the results of the modeling we can see the following:

- Both independent variables are valid and should be used in modeling
- The extent of tracking the original series and very high
- Durbin-Watson statistics indicate a possible serial correlation of the residuals. Accordingly, will be made a correlogram of the residuals to check if there is any dependence on any delay in the series

Table 11 gives the correlogram of the residuals of the model.

Table 11.Correlogram of the residuals of the model

Date: 04/29/18 Time: 18:26
Sample: 2010M01 2017M12
Included observations: 95
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.267	0.267	7.0058	
		2	-0.040	-0.120	7.1654	
		3	-0.176	-0.144	10.273	0.001
		4	-0.090	-0.005	11.089	0.004
		5	-0.088	-0.091	11.888	0.008
		6	-0.035	-0.022	12.014	0.017
		7	-0.010	-0.018	12.025	0.034
		8	0.006	-0.019	12.029	0.061
		9	0.041	0.031	12.211	0.094
		10	-0.009	-0.046	12.220	0.142
		11	0.097	0.120	13.242	0.152
		12	-0.056	-0.124	13.592	0.192

The correlogram shows that the correlation of the first delay goes beyond the confidence interval. This means that the model should be changed to obtain the validity of the modeling.

And other tested models did not produce much better results in modeling the series. For this reason, was tested a possible interruption of the series. The results of this test are given in Table 12.

Table 12.Test on structural break(s) of the series

Null Hypothesis: DOMASNI_JZPR_D has a unit root
Trend Specification: Intercept only
Break Specification: Intercept only
Break Type: Innovational outlier

Break Date: 2015M05
Break Selection: Minimize Dickey-Fuller t-statistic
Lag Length: 10 (Automatic - based on Schwarz information criterion, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-37.54798	< 0.01
Test critical values:	1% level	-4.949133	
	5% level	-4.443649	
	10% level	-4.193627	

From the results given in Table 12 it can be noted that the series has a structural change in 2015, which means that the series cannot be modeled, because there is not enough data after the structural change. For modeling this series, is needed a model that is resistant to structural change of the series.

CONCLUSION

Modeling time series in tourism can be very challenging, especially in the period of development of tourism in some destination or region. We can see from the shape of the series, but also from the parameters of the series and test of structural break that in both series we have structural break. For the series of foreign tourists we have enough data after the structural break, so we can create valid model and we can make forecasting of future values. In opposite, for the series of domestic tourists, we have structural break in middle of 2015, and we don't have enough data after the structural break, so we cannot create a model for this series. Some models, based on artificial neural networks are able to create model for time series with included structural break, but they need reach input at input layer. However, this series, according to prior behavior isn't that interesting as it is the series of foreign tourists. For the series of domestic tourists we have variation of the values during the last ten years.

REFERENCES

Andreeski, C., & Vasant, P. (2009). Comparative analysis on time series with included structural break. *Proceedings of the Second Global Conference on Power Control and Optimization* (pp. 217-224). Bali: American Institute of Physics.

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

Benkachca , S., Behra, J., & El Hassani, H. (August 2013). Causal Method and Time Series Forecasting model based on Artificial Neural Network. *International Journal of computer Applications (0975-8887)*, 37-42.

Chen, K. (2011). Combining linear and nonlinear model in forecasting tourism demand. In *Expert Systems with Applications* (pp. 10368-10376). Vol. 38.

Ette, E. H. (2012). A Seasonal Arima Model for Nigerian Gross Domestic Product. *Developing Country Studies, Vol 2, No. 3*, 50-62.

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

Oscar, C., Enric , M., & Salvador , T. (2013). *Tourism demand forecasting with different neural networks models.* Barcelona: Research Institute of Applied Economics.

Petrevska, Biljana;. (2015). Predicting tourism demand by A.R.I.M.A. models. *Economic Research, doi 10.1080/1331677X.2017.1314822*, 939-950.

Ruey-Chyn, T., & Ting-Chun, K. (2014). TOURISM DEMAND FORECASTING USING A NOVEL high precision fuzzy time series model. *International Journal of Innovative Computing, Information and Control*, 695-701.

Shengwei, W., Juan, F., & Gang, L. (2013). Application of seasonal time series model in the precipitation forecast. *Mathematical and Computer Modelling, Volume 58, Issues 3–4*, 677-683.