Forecasting Tourism Demand in Croatia:

A Comparison of Different Extrapolative Methods

Tea Baldigara[1]

[1] Faculty of Tourism and Hospitality Management, Opatija, Croatia

Correspondence: Tea Baldigara, Faculty of Tourism and Hospitality Management, Opatija, Croatia. Tel: 385-51-294-684. E-mail: Tea.Baldigara@fthm.hr

Received: March 27, 2013 Accepted: April 16, 2013 Online Published: April 23, 2013

doi:10.5430/jbar.v2n1p84 URL: http://dx.doi.org/10.5430/jbar.v2n1p84

Abstract

The paper investigates the forecasting accuracy of different basic extrapolative methods in modelling international tourism demand in Croatia. The study compares the results of five basic time-series forecasting methods used to predict foreign tourists' nights, namely the Naïve 2 trend, the double moving average with linear trend, the double exponential smoothing, the linear trend time and the autoregressive method. According to the diagnostic all used models show good forecasting performances, but the double moving average method performed the best forecasting performance due to the smallest value of the mean absolute percentage error.

Keywords: Forecasting, Foreign tourists' nights, Croatia, basic extrapolative methods, Forecasting accuracy

1. Introduction

Croatia is a significant Mediterranean tourist destination, and tourism is a significant source of profit for a wide range of activities. According to the Croatian Ministry of Tourism in 2012 there were 62.7 million of tourists nights registered and among these 57.5 million were realized by foreign tourists. Since 2009 the number of foreign tourist nights reveals aslightly upward trend with a growth rate of 4.5%.Foreign tourists account for a total of 85% of the total Croatian tourist activity. Table 1.gives an overview of the number of tourist nights according to the country of residence.

Table 1. Foreign tourists' activity In Croatia from 2009 to 2012

Year	Foreign tourist nights in 000				
	GERMANY	SLOVENIA	AUSTRIA	ITALY	CZECH REPUBLIC
2009	11 167	5513	4258	4925	3962
2010	11 476	5885	4420	4732	4170
2011	12 487	6398	4836	4994	4389
2012	13 947	6239	5104	4535	4519
Growth rate 2008–2011 (in %)	+7.8	+4.3	+6.3	-2.5	+4.5

Source: Ministry of Tourism: http://www.mint.hr/UserDocsImages/2012-turist-p-analiza.pdf (accessed 10 March 2013)

In 2012 most of the foreign tourists came from Germany and realized 13.9 million of nights, registering a growth rate of 7.8%. Among the top five countries by tourist nights, only Italy registered a negative growth rate of 2.5%. Foreign tourists' activity, expressed in tourist nights, registered in the past four years a growth rate of 6.3%. Due to the importance of international tourism demand for the Croatian tourism sector this study attempt to present and emphasize the necessity to implement quantitative methods, in modelling and forecasting tourism demand and all its components.

The study is structured in five sections. After the introductory section, some basic theoretical concepts of the time-series analysis are given and a brief literature review is provided. The third section deals with data and methodology. The theoretical concepts of forecasting techniques adopted in the study, and the essentials of forecasting errors for evaluating the forecasting accuracy are briefly described.In the fourth section the comparison of the selected extrapolative forecasting techniques is made and the obtained results are discussed. In the conclusion the results of the study are mentioned and some suggestions of the future research directions are given.

2. Theoretical Background and Literature Review

Along with the growing significance of the international tourism in Croatia, it should be a growing interest in modelling and forecasting tourism demand and its components. "Modelling tourism demand in order to analyse the effects of various determinants, and accurate forecasting of future tourism demand, are two of the major focuses of tourism demand studies" (Song & Witt, 2005). Accurate forecasts of tourism demand and its features can certainly improve planning and decision-making.

The primary goal of time-series analysis is finding out of an optimal model which will describe the dynamic system as well as forecasting of its future state on the basis of known present and past values. Time-series models have been widely used in the past for tourism demand forecasting. A time-series model explains a variable with regard to its own past and a random disturbance term. The aim of time-series analysis is to explore the historic trends and different patterns, such as seasonality, trend and cyclical patterns of the time-series involved and to try to predict the future of the series based on the trends and patterns identified in the model (Box & Jenkins, 1970).

It has always been difficult to model and forecast tourism demand because of its complexity. Fretchling (2001) states that "the nature of tourism demand presents a number of special challenges to the forecaster that does not afflict those in others industries". Among these challenges the lack of historical data, the volatility of tourism demand, its sensitiveness to catastrophic influences and the complexity of tourism behaviour can be mentioned. Nevertheless, forecasting plays a major role in tourism researching and the basic extrapolative forecasting methods applied in this study, may present a helpful tool to make more accurate future Croatian tourism demand forecasts.

In a comprehensive study Song and Li (2008) reviewed 121 papers on tourism demand modelling and forecasting published in the period between 2000 and 2007 and they found out that in 72 studies time-series models, were used to model or forecast tourism demand. Different versions of the exponential smoothing techniques were used in tourism demand forecasting and as a benchmark in forecasting accuracy comparison. Buger, Dohnal, Katharda and Law in a study (2001) used, among others, the exponential smoothing technique in forecasting the US demand for travel to Durban, South Africa. Hue at al. (2004) used the Naïve 2 and the exponential smoothing techniques in forecasting the number of restaurants daily customers. Law (2004) used the Naïve 2, the exponential smoothing and the trend extrapolation in forecasting hotel room occupancy rate. In the detailed outline of researches Song and Li (2008) listed 121 publications and among these there are 22 publications that used the Naïve 2, the double exponential smoothing, the moving average, the linear trend and the autoregressive method to compare forecasting accuracy. According to Song and Li (2008) "Naïve 2 (or constant-growth rate), exponential smoothing models and simple autoregressive models have frequently appeared in the post-2000 studies, but as in earlier tourism forecasting studies, they are usually used as benchmarks for forecasting accuracy evaluation." Those models are therefore used in this study to investigate the forecasting accuracy of the selected time-series forecasting methods.

3. Data and Methodology

According to Song, Witt and Li (2012) tourist arrivals is the most commonly used measure of international tourism demand, followed by tourist expenditure and tourist nights in registered accommodation. This study considers foreign tourist nights as a measure of international tourism demand in Croatia over the time period 1993-2012. Data were taken from the Croatian Central Bureau of Statistics. Figure 1.shows the actual data that are used in the forecasting model building process.

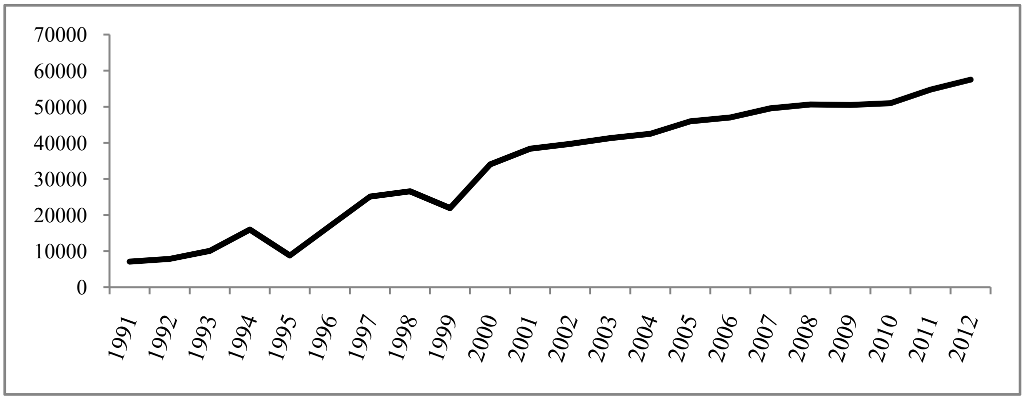

Figure 1. Foreign Tourist Nights in Croatia from 1993 to 2011

The plot of the foreign tourist nights illustrates an upward trend component in time. Therefore, specific forecasting methods, that take into account the trend component, should be used to forecasts future time-series values. Among the different forecasting model, built to capture the trend component, the Naïve linear trend model (the Naïve 2), the double moving average with linear trend technique, the double exponential smoothing with trend, the linear time trend method and the autoregressive method are used in this study. Figure 1.shows that data are available for a time period 1991-2012 (*n=22*). The historical data are used in the attempt to calibrate five extrapolative forecasting models in order to make predictions over the expected 2 future time periods. The forecasting models are estimated based on the sample data and, then forecasts for 2013 and 2014 are generated. Latter, the comparison of ex post forecasts between models is made in order to discuss the forecasting performances.

The following describes the theoretical concepts of the forecasting techniques adopted in this study and the essentials of forecasting errors for evaluating the forecasting accuracy.

3.1 The Naïve 2

When dealing with time-series that trends upward or downward the Naïve trend model (the "Naïve 2") can be used to forecast future values. It is calculated by the following expression:

$$F_{t+1} = y_t + (y_t - y_{t-1}) \tag{1}$$

where

F = forecast value

Y = actual value

t = some time period

It is a simple extrapolative forecasting technique based on the naïve approach that states that the differences of consecutive values are constant. The Naïve models are usually used as a benchmark forecast, namely for comparison with the forecast generated by other more sophisticated forecasting methods.

3.2 The Double Moving Average

The double moving average with linear trend (DMA) technique is a variation of the moving average procedure that better captures the presence of the linear trend component. It calculates a second moving average from the original moving average, using the same value for M. As soon as both single and double moving averages are available, these averages are used to compute a slope and a trend coefficient, namely the average change over *h* periods in order to make forecasts one or more periods ahead. The expression to calculate the forecast *h* periods into the future is:

$$Y_{t+h} = a_t + b_t h \tag{2}$$

where

a = intercept

b = slope coefficient

t = some time period

h = number of time periods ahead to be forecast

This technique is also called the Moving Average with Linear Trend for the forecast (Jaffe & McGee, 2000)

3.3 The Double Brown's Exponential Smoothing

The formulation of exponential smoothing forecasting methods arose in the 1950's form the original work of Brown and Holt and their basic ideas is to calibrate forecast future values as weighted averages of historical values, but with more recent values carrying more weight than the past observations. Second order or double exponential smoothing (DES) technique should be used when time-series data has only trend and no seasonality. Among the different developed DES model the simplest to apply is the Brown's one-parameter adaptive method (Fretchling, 2001). The following equations are used in double exponential smoothing with Brown's method (Fretchling, 2001):

$$\begin{aligned} Level&: \alpha A_t + (1 - A)(L_{t-1} + b_{t-1}) \\ Trend&: b_t = \alpha(L_t - L_{t-1}) + (1 - \alpha)b_{t-1} \\ Forecast&: F_{t+h} = L_t + hb_t \end{aligned} \tag{3}$$

where

L = level of the series

α = level and trend smoothing constant between 0 and 1

A = actual value

b = trend of the series

t = some time period

h = number of time periods ahead to be forecast

3.4 The Linear Time Regression

The linear time trend as a forecasting method regresses a time series as a measure of tourism demand on the periods to which is related to. The linear time trend regression indicates the number of units the dependent variable change for each unit change in time. The form of the simple regression is:

$$Y_t = a + bX_t + e_t \tag{4}$$

where

Y = the dependent variable

t = some time period

a = intercept constant

b = slope coefficient

X = time, explanatory variable

e = residual

3.5 The Autoregressive Method

In analysing a tourism time-series it is not rare to have a relationship between the current period's value and the previous or several previous periods values. Such relationships can be well calibrated by an autoregressive model, which regresses the current period's value on some chosen past values form the same time series. The mathematical expression of the autoregressive model is:

$$F_t = a + b_1A_{t-1} + b_2A_{t-2} + \cdots + b_nA_{t-n} \tag{5}$$

where

F = forecasted value

a = estimated constant

b = estimated coefficient

A = actual value of the time-series

t = some time period

n = number of past values included

3.6 Accuracy Forecast Measures

According to Fretchling (2001) “the most familiar concept of forecasting accuracy is called “error magnitude accuracy”, and relates to forecast error associated with a particular forecasting model.” The error magnitude is defined as:

$$e_t = A_t - F_t \tag{6}$$

where

t = some time period

e = forecast error

A = actual value of the variable being forecast

F = forecast value

Among the different methods developed to measure the error magnitude accuracy this study will consider the Mean Absolute Deviation (MAD), the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE), and the Tracking Signal (TS). The following describes the basic concepts of the different forecast accuracy measures used in this paper.

The Absolute Deviation (MAD) is the most popular and simplest to use measure of forecasting accuracy. It is the average of the difference between the forecast and the actual value and it is measured in the same units as the original data. The expression to calculate MAD is:

$$MAD = \frac{1}{n}\sum_{t=1}^{n}|(A_t - F_t)| \tag{7}$$

The smaller the value of MAD is, the more accurate is the forecast.

The Root Mean Square Error (RMSE) is computed by the following expression:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2} \tag{8}$$

For purpose of communicating results, it is usually best to report the Root Mean Square Error (RMSE) rather than MSE, because the RMSE is measured in the same units as the data, rather than in squared units, and is representative of a “typical” error (Nau, 2013). The RMSE is usually more sensitive than other forecast accuracy measures. RMSE and MAD can only be compared between models whose errors are measured in the same units.

The Mean Absolute Percentage Error (MAPE) is expressed in generic percentage terms and it is computed by the following formula:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|(A_t - F_t)|}{A_t} \cdot 100 \tag{9}$$

MAPE is a simple measure that permits to compare the accuracy of different models, with different time periods and numbers of observations.

According to Baggio and Klobas (2011) “a rough scale for the accuracy of a model can be based on MAPE” following the suggestions gave in the table below.

Table 2. MAPE and forecasting accuracy

MAPE	Forecasting accuracy
Less than 10%	Highly accurate
10-20%	Good
20-50%	Reasonable
Greater than 50%	Inaccurate

Source: Baggio, R. & Klobass, J. (2011)

The Tracking Signal (TS) monitors any forecasts that have been made in comparison with actual values, and warns when there are unexpected deviations of the values from the forecasts. The tracking signal is a simple indicator that forecast bias is present in the forecast model. The TS is a measure used to evaluate if the actual values do not reflect the assumptions in the forecasts. The TS can be calculated using the following formula:

$$TS = \frac{\sum_{i=1}^{n}(A_t - F_t)}{MAD} \tag{10}$$

The TS to worn if there is a persistent tendency for actual values to be higher or lower systematically. If forecasted values are consistently lower than the actual values, then there is persistent under forecasting and the TS will be positive. The TS should pass a threshold test to be significant. According to Fretchling (2001) if the data are well behaved the threshold can be set at – 3 and +3, on the other hand if there are some difficulties fitting a good forecasting model to the actual data series the threshold may be set at – 5 to +5. As long as the tracking signal is between –4 and + 4, one can assume the forecasting model is working correctly (Bozarth, 2011).If the TS is greater than +4 then there is persistent under-forecasting. On the other hand, if this is less than – 4 then, there is persistent over-forecasting.

4. Results and Discussion

In this section the comparison of the selected techniques in forecasting international tourism demand in Croatia, expressed in foreign tourist nights', is made and the obtained results are discussed.

The double moving average (3x3), used for additional smoothing of the single moving average methods, were used to model the original time-series. The double exponential smoothing with linear trend was used because of the upward trend present in the series. The parameter α smoothed the level equation and its value was set up at 0.48. In the linear time regression the changes in time intervals, years, as the independent variable, were used to explain the variations in the forecasted foreign tourism nights' time series. The values of the intercept and the slope coefficient were compute using the ordinary least square methods, which is designed to make the sum of the squared errors as small as possible. The following trend time regression was computed:

$$Y_t = 4362.13 + 2557.641X_t + e_t \tag{11}$$

The value of the determination coefficient is high (R^2=0.95328) and shows that 95% variations in the number of foreign tourist nights are explained by the variations in time.

In modelling the foreign tourist nights time-series with the autoregressive method, among different possible combinations of the time-series lagged, the model with three lagged values only indicated that all the estimated coefficient were significantly different from zero. The chosen model produced also the largest coefficient of determination and the lowest measures of error magnitude. In the chosen autoregressive model the explanatory variable is represented by the third lagged values and its states that the regressed values can be a good basis for forecasting the empirical time-series.

The results of the calculus are shown in Figure 2., which illustrates the results of the selected model by comparing the actual data and the different forecasts obtained.

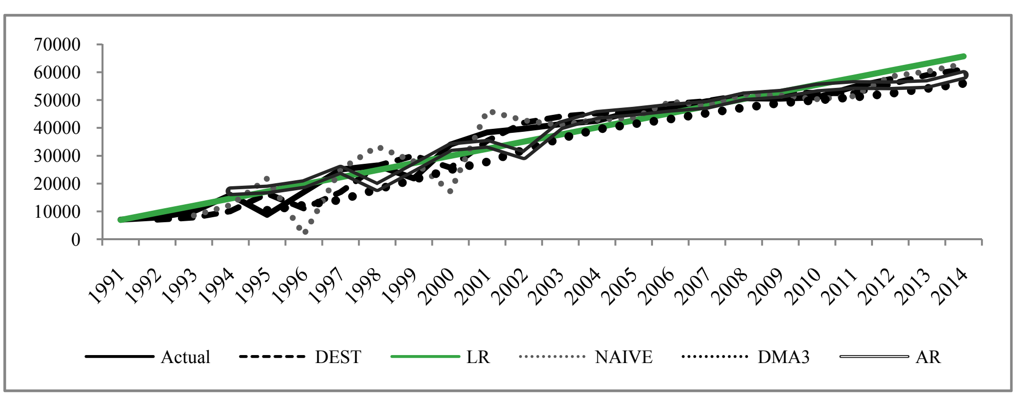

Figure 2. Forecast of foreign tourist nights: comparison of forecasting models

For the purpose of testing the presence of autocorrelation, the residuals were plotted versus time. The obtained results are shown in the figure below.

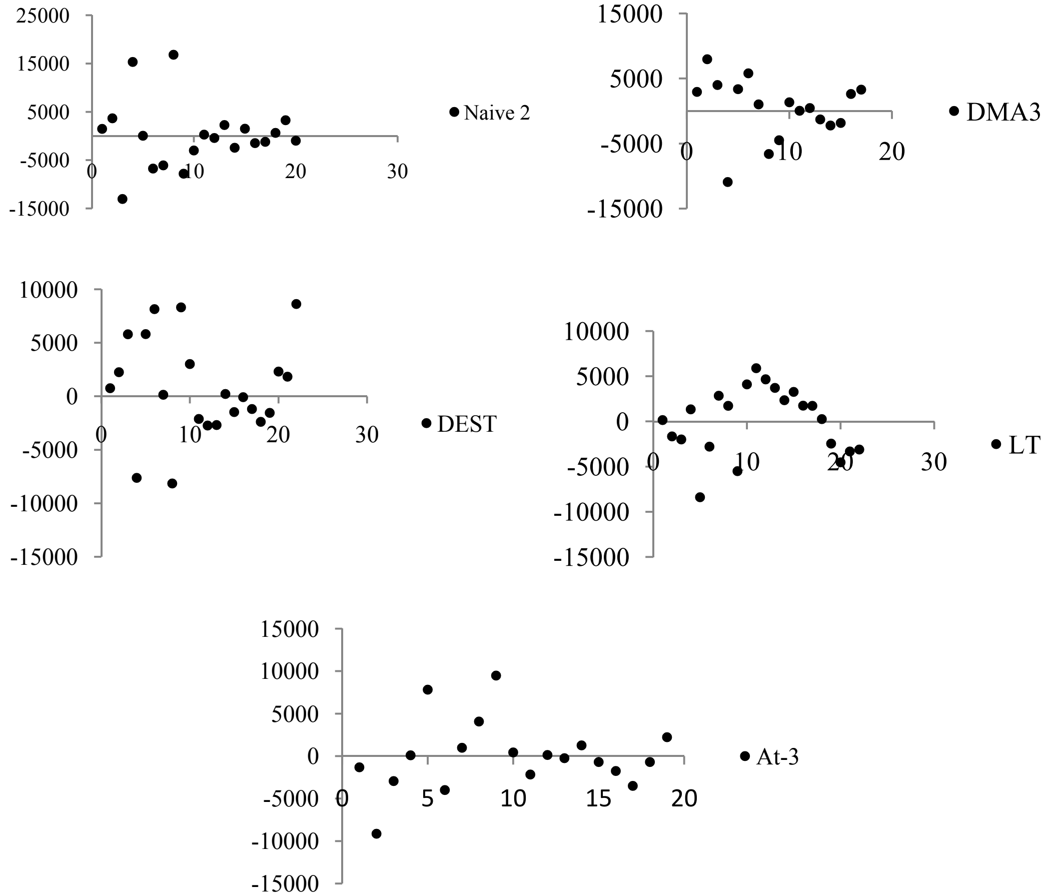

Figure 3. Plots of the residuals versus time

The plot of the LT model and the computed DW test statistics (DW=1.208) reveal the presence of positive autocorrelation. The method of generalized least squares was therefore used to eliminate the autocorrelation in the residuals and the generalized time regression was computed:

$$Y_t^* = 2871.744 + 2534.895X_t^* + v_t \tag{12}$$

The DW test statistics was found to be 1.9 and the null-hypothesis of no autcorrelation can be accepted. The slope coefficient indicates that the number of foreign tourist nights' increases of 2534.895 for each for one unit change in year. The value of the determination coefficient of 0.890605 indicates that 89% of the variations in the number of foreign tourist nights' are explained by variations in time, as the explanatory variable.

When comparing the forecasting accuracy of the selected extrapolative techniques, the diagnostic reveals that all the used techniques, in term of analysed forecasting errors measures, reveals a good accuracy in modelling the original time-series. Table 3.gives the forecasting errors for all four techniques used. The results are obtained by computing the expressions explained earlier in this study.

Table 3. Errors of the selected forecasting techniques

ERRORS	FORECASTING TECHNIQUES				
	NAÏVE 2	DMA (3X3)	DEST (α=0.48)	LT	AR (A_{t-3})
MAD	4429.4	3546.725	3266.6	2616.618	2786.69
RMSE	6656	4521.246	4264.826	3394.542	4013.515
MAPE	21.86%	11.38%	15.81%	26.74%	12.72%
TS	0.4559	1.576634	2.6411	0.000	0.000
				R^2=0.890605	R^2=0.91
DW	2.9	1.8	2.28	1.9	1.63

The MAD values, as the absolute size of the errors, indicate that the models tend to over-forecast. The MAD and the RMS are both measured in foreign tourist nights', and it can be observed that the MAD values are slightly smaller the RMS, which is usual. There is no absolute criterion for a "good" value of RMSE of MAD; it depends on the units in which the variable is measured and on the degree of forecasting accuracy, as measured in those units, which is sought in a particular application (Nau, 2013). The smallest mean absolute error is performed by the LT model. The DMA model presents the smallest MAPE value, i.e. 11.38%. This value indicates a good forecasting accuracy. Only the naïve 2 model and the LT have a MAPE value greater than 20% which indicates a reasonable forecasting accuracy. The TS of all four compared model are within the interval of – 4 to +4, which means that the models can be used in forecasting the time-series. The mean forecast error, using the linear time regression is 0 and it shows a good forecasting accuracy, and for the other tested models it is greater than 0, which indicates that the models tend to under-forecast the actual time-series. Although the diagnostic for the selected models reveals that the four models do not significantly differ, it can be conclude that, the double moving average technique performed a good forecasting (smallest MAPE value) of the foreign tourist night in Croatia. According to Burger (Burger, 2001), "a tourist forecaster or policy maker can at least make a good estimate of visitor arrivals in the absence of structural data by just using the available time series data".

5. Conclusion

The aim of the presented study was to point out the necessity of using quantitative forecasting methods in analysing the Croatian tourism demand in all its aspects. By examining five basic forecasting time-series methods, namely the Naïve 2, the double exponential smoothing with trend, the double moving average, the linear time trend and the autoregressive method, the number of foreign tourist nights over the period 1991-2012 was used in an attempt to create models and predict the expected number of foreign tourist nights over two future periods. The research results shown that, the double moving average performed the best forecasting of foreign tourist nights for the tested period. Although, the forecasting models used in this study are well-known and often used in making forecasts, the author wanted to point out the necessity of more systematic quantitative researches of the Croatian tourism demand. In future, more sophisticated quantitative, both extrapolative and causal forecasting techniques, should be applied in investigating, modelling and forecasting tourism demand in Croatia, especially international, more intensively. The study wants to emphasize that, given the importance of tourism for Croatia's economic development, there is a lack of quantitative approaches to tourism demand modelling and forecasting. Such researches should be used as starting points for all those stakeholders involved in the tourism sector. The results of such studies and researches should offer recommendations for action to the national economic policy, evaluations of the effects and measured tourism demand components over long-term period and the possibility to make detailed analysis of the strengths and weaknesses of the Croatian tourism sector in order to reroute and develop this significant source of profit for the entire national economy. Profoundly aware that modelling tourism demand is a challenging and controversial issue, that the adequacy of a forecasting model is valued according to its out-of- sample forecasts, and that is still difficult to indicate which model or class of model is more adequate for tourism demand modelling, the author of this study wished to highlight the necessity of more systematic quantitative analysis of Croatia's tourism demand in all its determinants.

Acknowledgments

The author would like to express its thanks to the Faculty of Tourism and Hospitality Management in Opatija, University of Rijeka for providing facilities, as well as moral and financial support for this research.

References

Baggio, R. & Klobas, J. (2011). *Quantitative Methods in Tourism*, Bristol: Channel View Publications.

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Bozarth, C. (2013). Measuring Forecast Accuracy: Approaches to Forecasting : A Tutorial. Retreived February 2, 2013, from http://scm.ncsu.edu/scm-articles/article/measuring-forecast-accuracy-approaches-to-forecasting-a-tutorial.

Burger, C.J.S.C. (2011). A practitioners guide to time – series methods for tourism demand forecasting – a case study of Durban, South Africa, *Tourism Management, 22*, 403-409. http://dx.doi.org/10.1016/S0261-5177(00)00068-6.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, *24(3),* 323-330. http://dx.doi.org/10.1016/S0261-5177(02)00068-7

Frechtling, D.C. (2001). *Forecasting Tourism Demand: Methods and Strategies.* Woburn: Butterworth Heinemann, Elseivierplc group.

Gerard S., & Dharmaratne, G.S. (1995). Forecasting tourists' arrivals in Barbados. *Annals of Tourism Research*, *22(4),* 804-818. http://dx.doi.org/10.1016/0160-7383(95)00022-3.

Jaffe, R.A., & McGee, M. (2000). An Introduction to Time Series Analysis and Forecasting: With Application of Sas and Spss. New York: Academic Press.

Lee, C.K., Song, H.J., & Mjelde, J.W. (2008). The forecasting of International Expo tourism using quantitative and qualitative techniques. *Tourism Management*, *29 (6)*, 1084-1098. http://dx.doi.org/10.1016/j.tourman.2008.02.007

Lim, C., & McAleer, M. (2001). *Time series forecasts of international tourism demand to Australia.* Discussion paper *533,* Osaka: The Institute of Social and Economic Research Osaka University.

Lin, C., Chen, H. & Lee, T. (2011). Forecasting Tourism Demand Using Time Series, Artificial Neural Networks and Multivariate Adaptive Regression Splines: Evidence from Taiwan. *International Journal of Business Administration (2) 2*, 14-24. http://dx.doi.org/10.5430/ijba.v2n2p14

Nau, R.F. (2013). What's the bottom line? How to compare models. Retrieved March 1, 2013, from http://people.duke.edu/~rnau/compare.htm

Shen, S., Li, G., & Song, H. (2011). Combination Forecasts of International Tourism Demand. *Annals of Tourism Research*, *(38)1,* 72-89. http://dx.doi.org/10.1016/j.annals.2010.05.003.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting – A review of recent research. *Tourism Management*, *29*, 203-220. http://dx.doi.org/10.1016/j.tourman.2007.07.016

Song, H., Witt, S., & Gang, L. (2012). *The Advanced Econometrics of Tourism Demand*. London: Routledge.