PLANNING AND ESTIMATING TOURISM DEMAND - THE CASE OF MACEDONIA

Abstract

This paper underlines the importance of planning process and estimation of tourism demand which is a foundation on which all tourism-related business decisions ultimately rest. Due to the fact that the concept of tourism planning could not be applied if forecasting tourism demand is neglected, the paper provides a medium-term estimation of foreign tourism demand for Macedonian destinations for the period ending by 2014. To this end, the Box-Jenkins methodology was used as one of the quantitative methods commonly applied in estimations. The paper brings out several alternative specifications in modelling the original time series. Despite the fact that this method is not capable of explaining the driving factors behind these results, the estimated values can serve as a solid base in the preparation of tourism development plan. Additionally, the paper refers to the importance of applying forecasting methods, thus being a pioneer in this issue when addressing the tourism industry in Macedonia.

Key words: Tourism planning; Estimation; Box-Jenkins methodology; Tourism demand; Macedonia.

INTRODUCTION

In the past few decades, tourism has emerged as one of the major industries in the world economy, by benefiting transportation, accommodation, catering, entertainment, retailing and many other sectors. Each country attempts to develop tourism industry and increase the number of incoming visitors for several reasons. It affects aggregated demand, domestic output and employment, it is a source of economic growth, and influences the balance of payments. Consequently, the contribution of worldwide tourism industry to the global economic development is significantly important. In such cases, the governments should pay much attention to the growth of the number of tourists into the country. So, planning and estimating tourism demand becomes highly important. With an appropriate forecasting model that could validly predict the tourism demand, the government would be able to plan properly and effectively tourism development in order to choose an appropriate strategy for its economic welfare. A reliable estimation is needed and as such, plays a major role in formulating and implementing appropriate medium to long term tourism strategies. Therefore, accurate estimations of tourism demand are essential for efficient tourism planning.

This paper provides a medium-term estimation of foreign tourism demand for Macedonian destinations for the period ending by 2014. In this respect, the projected values may serve as a background for preparation of tourism development plan in Macedonia. Moreover, the paper makes an attempt to introducing econometric modelling for the first time in academic research in Macedonia and refers to the importance of their applications.
TOURISM PLANNING vs. ESTIMATING TOURISM DEMAND

There is an obvious relationship between the concept of tourism planning and estimating tourism trends. Namely, estimation permits planners and policy-makers to reach decisions before the occurrence of the events. Without reliable estimates of future demand, it is difficult, if not impossible, to formulate adequate tourism development plan and policy (Vanhove, 1978). It should be noted that the main principles must always prevail, in order the tourism policy to ensure that visitors are hosted in a way which maximizes the benefits to stakeholders, while minimizing the negative effects, costs, and impacts associated with ensuring the success of the destination (Goeldner and Ritchie, 2006). However, all efforts in order to considerate and understand the interrelated nature of tourism industry require monitoring and evaluation when tourism policy issues are involved (Edgell et al., 2008).

Additionally, the factors which can influence tourism demand are normally to be found within the tourist-generating countries (Lickorish and Jenkins, 1997), but also may initiate from all sectors of the economy - individuals and households, private businesses and the public sector (Sinclair and Stabler, 1997).

Estimating tourism demand has attracted a lot of attention by both, the academic literature and tourism practitioners (Song and Turner, 2006). It is more than obvious that the success of many businesses depends largely or totally on the state of tourism demand. More precisely, the demand is a key determinant of business profitability and its estimations constitute a very important element in the whole planning process. Estimation of tourism demand can be helpful to economic planners in reducing the risk of decisions regarding the future (Frechtling, 2001). In the same line, it is important to the tourism manager and to those who depend on that manager, since more accurate estimations reduce the risks of decisions more than do less accurate ones. Hence, the accuracy is one of the most important forecast evaluation criterions (Witt and Witt, 1992). It is obvious that a wide range of techniques and procedures available for tourism policy analysis must be introduced in order fulfill tourism planning in adequate manner (Chowdhury and Kirkpatrick, 1994). Besides, estimating can serve as a mean to deal with the alternative future although it may evolve in strikingly different ways (Coates and Jarratt, 1989). Anyway, anticipating tourism flows considers the historical facts as well as the scientific knowledge in order to create images of what may happen in future (Cornish, 1977) because only then, the forecasting process may allow the prediction of future.

METHODS FOR ESTIMATING TOURISM DEMAND

There is a large body of literature regarding application of methods for estimating tourism demand. Namely, numerous researchers have been involved and a wide variety of techniques has been used. In principle, all methods are generally categorized in two-categories: qualitative and quantitative (Song and Li, 2008). The qualitative methods use pooled opinions of experts to organize the past information of the variable and often are recommended as methods which seldom generate better predictions (Hall, 2005). The quantitative methods organize past information about a phenomenon by mathematical rules and assume that at least some elements of past patterns will continue into the future (Makridakis et al., 1998).

Due to the seasonal character of tourism, numerous studies attempt to account it in the estimations, like: Diebold and Kilian, 2000; Gustavsson and Nordstrom, 2001; Turner and
Witt, 2001; Lim and McAleer, 2001; Kulendran and Witt, 2003; Rodrigues and Gouveia, 2004; Kulendran and Wong, 2005; Alleyne, 2006; Coshall, 2006; Lee et al., 2008. Generally, the standard econometric techniques based on unit roots or seasonal unit root test statistics are used (Dickey and Fuller, 1979; Dickey et al., 1984; Phillips and Perron, 1988; Hylleberg et al., 1990; Kwiatkowski et al., 1992).

Regardless the method, it is expected that the final model chosen for estimations will produce projections which are as precise as possible. However, it is not always the case due to lack of sufficient time series data, measurement errors, or even, unclear picture for the system of tourism demand (Song and Witt, 2000). So, certain evaluation criteria are used in order to select potential starting methods, as well as to identify an adequate model. However, no individual model consistently performs well in all situations (Witt and Song, 2002) meaning that no single forecasting model is the best for all situations under all circumstances (Makridakis et al., 1982). Therefore, solution is seen in proposing combination models since one cannot identify the true process exactly, but combining often results in a prediction accuracy which is higher than the one of the individual models (Lawrence et al., 1986; Makridakis, 1989; Makridakis and Winkler, 1983). Furthermore, the performance of the forecasting models varies according to the length of the forecasting horizons (Li et al., 2005). Forecasting domestic tourist flows is considerably easier than forecasting international tourist flows’ over a one-year horizon (Witt et al., 1992).

Tourism demand can be expressed in a variety of way. Some explained it by consumer expenditure or receipts (Grouch, 1992; Li et al., 2004) as the only one applicable variable which can be directly translated into economic impact (Sheldon, 1993). Others employed tourist expenditure on particular tourism product categories, such as meal expenditure (Au and Law, 2002), sightseeing expenditure and expenditure (Au and Law, 2000). On the other hand, others made their focus on tourist typologies, motivation, determinants of choice of activities and demand (Johnson and Thomas, 1992). Even more, tourism demand can be measured by visitors’ use of a good or service (Frechtling, 2001), tourism revenues (Akal, 2004), tourism employment (Witt et al., 2004) and tourism import and export (Smeral, 2004). However, the tourist arrivals variable is the most popular measurement of tourism demand (Crouch, 1994). This variable further may be decomposed into holiday tourist arrivals, business tourist arrivals, tourist arrivals for visiting friends and relatives purposes (Turner and Witt, 2001a, 2001b; Kulendran and Wong, 2005), and tourist arrivals by air (Coshall, 2005; Rosselló, 2001).

**METHODOLOGY**

In this section we introduce the Box-Jenkins methodology (Box and Jenkins, 1976) as one of the quantitative methods commonly applied in estimating, known as autoregressive integrated moving averages (ARIMA) models. It is the most popular linear model for forecasting time series and enjoys great success in academic research (Qu and Zhang, 1996; Law, 2000 and 2004; Goh and Law, 2002; Kulendran and Shan, 2002; Huang and Min, 2002; Lim and McAleer, 2002; Coshall, 2005).

We model the original time series - the number of foreign tourists in Macedonia in the period 1956–2008 (State Statistical Office, 2008 and 2009), hence the sample consists of 53 observations. Further on, four alternative specifications are used to model the original series: ARIMA(1.1.1) with dummy, ARIMA(2.1.2), restricted ARIMA(1.1.10) with dummy, and restricted ARIMA(10.1.10).

Given that the basic assumption for applying the Box-Jenkins methodology is obtaining stationarity of the time series, the first step in the analysis is to perform the
stationarity test. In that respect, the correlogram of the series is used. The results emphasizes that the series is non-stationary, because for stationary series the autocorrelation coefficients of all pairs of observations are equal to zero. If dealing with a random process, then the autocorrelation coefficients are approximately characterized by the normal distribution, with a zero mean and variance of 1/n, where n is the sample size (Gujarati, 1995). So, the statistical significance of the calculated autocorrelation coefficients is set by formula for standard error, thus resulting as: \( \sqrt{1/n} = 0.137 \). According to the table for normal distribution, we can calculate the 95% confidence interval for the autocorrelation coefficients:
Confidence interval = \( \pm 1.96 \times 0.137 = \pm 0.269 \).

If the calculated autocorrelation coefficient is within the confidence interval, it means that the null hypothesis that the true autocorrelation coefficient of the population is zero \( (H_0: \rho_k = 0) \), cannot be rejected. The large number of statistically significant coefficients confirms that the series is non-stationary.

However, given the problems with individual testing of the significance of autocorrelation coefficients, the joint hypothesis that all autocorrelation coefficients are equal to zero, is tested. This test is usually made with Ljung-Box statistic - LB, calculated with the formula:

\[
LB = n(n+2) \sum_{i=1}^{k} \frac{\rho_i^2}{n-i}
\]  

The LB-statistics tests the null hypothesis where there is no autocorrelation for all coefficients at certain number of time lags. Further on, if the null hypothesis is true, the LB-statistics asymptotically follows the \( \chi^2 \) distribution with degrees of freedom equal to the number of autocorrelation coefficients. In our calculations, the LB-statistics by far exceeds the critical values. So, this test shows that the null hypothesis can be rejected, which by all means is a proof that the analysed time series is non-stationary. Yet, it is known that the LB-statistics has low power, because the significant coefficients can be neutralised by the insignificant ones. Hence, the evidence gained by the LB-statistics is additionally tested by employing two unit root tests: the Augmented Dickey-Fuller - ADF (Dickey and Fuller, 1979) and the Phillips-Perron test - PP (Phillips and Peron, 1988).

**Table 1. Stationarity tests for the number of foreign tourists in Macedonia**

<table>
<thead>
<tr>
<th>Test</th>
<th>constant</th>
<th>constant + trend</th>
<th>none</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF</td>
<td>-1.547875 (0.5016)</td>
<td>-1.498094 (0.8174)</td>
<td>-0.511774 (0.4899)</td>
</tr>
<tr>
<td>PP</td>
<td>-1.599661 (0.4756)</td>
<td>-1.496664 (0.8182)</td>
<td>-0.557843 (0.4708)</td>
</tr>
</tbody>
</table>

In the first row of Table 1, the values of the ADF-test are shown in its three variants, and in all cases, the null hypothesis for the presence of unit root, cannot be rejected. Consequently, this test suggests that the series is non-stationary. However, due to the fact that the ADF-test has low power, the results are checked with the PP-test. As shown in the second row of Table 1, all the variants of the PP-test show that the null hypothesis of a unit root cannot be rejected. Hence, this test, too, suggests that the series is non-stationary. As mentioned previously, if the time series is non-stationary, than the Box-Jenkins methodology cannot be applied. It means that it is necessary to transform the series in order to make it stationary, which is done by differencing the original series.
When the series is differenced, one cannot observe some regular movement of the autocorrelation coefficients, which begin with low values, decreasing quickly to zero, and then moving in a wave-style, i.e. increasing and decreasing (Table 2). Also, one can observe the great value of the autocorrelation coefficient at lag 10. It can be explained with the fact that the series declines sharply twice with an interval of 10 years (the collapse of Yugoslavia in 1991, and the war in Macedonia, in 2001).

Table 2. Correlogram of the number of the foreign tourists in Macedonia (First Differences: 1956-2008)

<table>
<thead>
<tr>
<th>Lags ((k))</th>
<th>ACF ((\rho_k))</th>
<th>PCF ((\rho_{kk}))</th>
<th>LB-stat.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.256</td>
<td>0.256</td>
<td>3.5980</td>
<td>0.058</td>
</tr>
<tr>
<td>2</td>
<td>0.208</td>
<td>0.153</td>
<td>6.0355</td>
<td>0.049</td>
</tr>
<tr>
<td>3</td>
<td>0.150</td>
<td>0.073</td>
<td>7.3325</td>
<td>0.062</td>
</tr>
<tr>
<td>4</td>
<td>0.078</td>
<td>-0.000</td>
<td>7.6874</td>
<td>0.104</td>
</tr>
<tr>
<td>5</td>
<td>0.100</td>
<td>0.053</td>
<td>8.2802</td>
<td>0.141</td>
</tr>
<tr>
<td>6</td>
<td>-0.059</td>
<td>-0.123</td>
<td>8.4935</td>
<td>0.204</td>
</tr>
<tr>
<td>7</td>
<td>-0.110</td>
<td>-0.114</td>
<td>9.2448</td>
<td>0.236</td>
</tr>
<tr>
<td>8</td>
<td>0.024</td>
<td>0.093</td>
<td>9.2823</td>
<td>0.319</td>
</tr>
<tr>
<td>9</td>
<td>0.047</td>
<td>0.086</td>
<td>9.4246</td>
<td>0.399</td>
</tr>
<tr>
<td>10</td>
<td>0.199</td>
<td>0.207</td>
<td>12.078</td>
<td>0.280</td>
</tr>
<tr>
<td>11</td>
<td>-0.122</td>
<td>-0.247</td>
<td>13.103</td>
<td>0.287</td>
</tr>
<tr>
<td>12</td>
<td>-0.044</td>
<td>-0.038</td>
<td>13.242</td>
<td>0.352</td>
</tr>
<tr>
<td>13</td>
<td>0.026</td>
<td>-0.162</td>
<td>13.290</td>
<td>0.426</td>
</tr>
<tr>
<td>14</td>
<td>-0.168</td>
<td>0.005</td>
<td>15.375</td>
<td>0.353</td>
</tr>
<tr>
<td>15</td>
<td>-0.091</td>
<td>0.128</td>
<td>16.002</td>
<td>0.382</td>
</tr>
<tr>
<td>16</td>
<td>0.080</td>
<td>0.000</td>
<td>21.283</td>
<td>0.168</td>
</tr>
<tr>
<td>17</td>
<td>0.170</td>
<td>23.602</td>
<td>0.131</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-0.051</td>
<td>-0.001</td>
<td>23.819</td>
<td>0.161</td>
</tr>
</tbody>
</table>

In order to check the stationarity of the differenced series, the autocorrelation coefficients are individually tested with the confidence interval, which in this case is ± 0.272. Further on, it was shown that the null hypothesis that the true autocorrelation coefficients of the population are equal to zero cannot be rejected. Namely, the value of LB-statistic with 18 degrees of freedom is 23.819, which is not sufficient to reject the null. By all means, the above results show that by differencing of the original time series, stationarity is obtained. Yet, once again, in order to verify the results, the ADF-test and the PP-test are used. From Table 3, it can be concluded that the values of the statistics are highly significant, so once again, we can conclude that the differenced series is stationary.

Table 3. Stationarity tests of number of foreign tourists in Macedonia (First Differences: 1956-2008)

<table>
<thead>
<tr>
<th>Test</th>
<th>constant</th>
<th>constant + trend</th>
<th>none</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF</td>
<td>-5.376144</td>
<td>-5.445010</td>
<td>-5.415973</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0002)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>PP</td>
<td>-5.466517</td>
<td>-5.529348</td>
<td>-5.503297</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0002)</td>
<td>(0.0000)</td>
</tr>
</tbody>
</table>

After performing the additional tests, it can be concluded the Box-Jenkins methodology can be applied. The first step is to identify the appropriate model which will explain the time series movement. Here, crucial instruments are the sample autocorrelation - ACF and partial autocorrelation - PACF functions. The detailed analysis of both functions did not show any regularity in the movement of the autocorrelation coefficients (slow decay,
sharp picks at certain lags etc.), from which, the model could be identified. It is only certain that we have a mixed process, i.e. combination of autoregressive - AR and moving average - MA processes.

**ANALYSIS, DATA AND RESULTS**

Given the unclear character of the time series, we use four alternative specifications to model the original series: ARIMA(1.1.1) with dummy, ARIMA(2.1.2), restricted ARIMA(1.1.10) with dummy, and restricted ARIMA(10.1.10). All models represent the original time series in an adequate manner.

The ARIMA(2.1.2) model has a slightly high coefficient of determination, but the second MA is marginally insignificant at 5%. Also, the inverted MA root is 1, which makes the process inappropriate for estimation.

The restricted ARIMA(10.1.10) model tracks the original time series quite well, both terms are highly significant and the coefficient of determination is twice higher comparing to the previous model. However, the reciprocal root of the MA term is very near to 1. Yet, the statistical significance of the AR and MA terms at 10 lags is a consequence solely of the effects of the structural breaks in 1991 and in 2001. Because there is no reason for these events to take place in future on regular basis (in the time interval of 10 years), the inclusion of these AR and MA terms will not ensure adequate estimating in the future.

According to the statistical features of the models, two specifications out of four, show best results: the ARIMA(1.1.1) with dummy and the restricted ARIMA(1.1.10) with dummy. These models have the highest coefficients of determination and, also, they are favored on the basis of both the Akaike and the Schwarz information criteria. Further on, here, there are no problems with the inverted AR and MA roots. Yet, despite the positive statistical characteristics, the restricted ARIMA(1.1.10) with dummy is discarded due to the problems with the interpretation of the MA term. Once again, we emphasize that the inclusion of the MA term with a time lag of 10 periods ensures a good approximation of the time series in the past, but not in the future.

Hence, only the results of the ARIMA(1.1.1) with a dummy are presented here, as the most appropriate model for estimating the original time series.

**Table 4. ARIMA(1.1.1) model of the number of foreign tourists in Macedonia**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUMMY</td>
<td>-191192.4</td>
<td>21341.93</td>
<td>-8.958533</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.787363</td>
<td>0.165950</td>
<td>4.744591</td>
<td>0.0000</td>
</tr>
<tr>
<td>MA(1)</td>
<td>-0.423157</td>
<td>0.241562</td>
<td>-1.751749</td>
<td>0.0862</td>
</tr>
</tbody>
</table>

| R²       | 0.650544    | Akaike info criterion | 23.66973 |
| Adjusted R² | 0.635984 | Schwarz criterion   | 23.78337 |
| S.E. of regression | 32448.72 | Durbin-Watson stat | 2.089552 |
| Inverted AR roots | 0.79 |
| Inverted MA roots | 0.42 |
From Table 4, it can be concluded that the AR term is highly significant with value 0.8, which suggests a high level of persistence in the series. The second term is not significant at the level of 5%, but having in mind the relatively small sample, we decided to work with the model, because of its significance at 10%. In the same line, the coefficient before the dummy is highly significant. The adjusted R² is satisfactory high (0.64) having in mind that we have modeled the first difference of the series. The values of the inverted roots of the AR and MA terms lie within the unit root, which, once again, confirms that the chosen model is appropriate. Finally, as stated above, according to the information criteria, this model has better performances comparing to the previous ones.

**Chart 1. ARIMA(1.1.1) model of the number of foreign tourists in Macedonia**

![Chart 1](image)

Chart 1 shows the movement of the original time series together with the forecasted values obtained with the model. It is noticeable that the model follows the time series movement exceptionally well, which is supported by the within-sample forecasts. Namely, the average percentage error of the forecasts is 9.26%, and the Theil inequality coefficient is 0.043, with the Bias proportion of only 0.017, Variance proportion of 0.029 and high Covariance proportion of 0.954. The good performances of the chosen model allows for its application in estimating tourism demand in the future. Therefore, we project the number of foreign tourists for the period during 2009 – 2014 (Table 5).

**Table 5. Forecasting by ARIMA(1.1.1)**

<table>
<thead>
<tr>
<th>Year</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrivals</td>
<td>269 897</td>
<td>281 660</td>
<td>290 922</td>
<td>298 214</td>
<td>303 956</td>
<td>308 477</td>
</tr>
</tbody>
</table>

The results of the dynamic forecasts of the number of foreign tourists using ARIMA(1.1.1) with a dummy, point out that in the period 2009-2014, the number of foreign tourists will increase for about 10 000 tourists in the first years, and then a moderate growth can be expected, leading to the forecast of 308 477 foreign tourists in 2014. Although the projections obtained by the Box-Jenkins methodology cannot explain the factors behind these trends, they can serve as a solid base for the preparation of tourism development plan in Macedonia.

**CONCLUDING REMARKS**

The authors proposed to use ARIMA model to predict tourism demand in Macedonia, as one of the quantitative methods commonly applied in the estimation of tourism demand. While modeling the original time series, four alternative specifications were used. The result was a medium-run estimation of foreign tourism demand for Macedonian destinations. On the basis of the results from the dynamic forecasts for the period 2009-2014, the study found an
expectation of 25% increase of foreign tourists. Additionally, the results of the performance criteria showed that this model is credible and accurate for a medium term horizon. So, all key actors in tourism industry in Macedonia may have confidence to use it. Based on the empirical findings, this paper noted promising estimated values for future tourism trends.

Furthermore, the paper explained that the implemented model does not provide the solution, but only assists in finding it. Therefore, the estimated values may be a starting spot for inspiring interesting thoughts or, in some cases, may assist in furnishing additional points of indentifying a decision. This paper underlined the crucial role of estimation in tourism planning process. Moreover, the demand forecasts of foreign tourism might be the fundamental input for national regional and local development plans in order to reduce the risks. Furthermore, estimating tourism demand in Macedonia is a key factor that determines its competitiveness as tourist destination. Even though the model results are essential elements in the preparation of well-coordinated policies, they cannot do the job all by themselves. So, the outcomes showed that the Box-Jenkins methodology may serve as a framework, while the rest needs to be fulfilled with a lot of common sense and knowledge of details. To strengthen these results, a similar study should be replicated.

REFERENCES


