Forecasting Tourism demand in Cyprus

Maria Demetriadou

June 25, 2018

Abstract

The aim of this essay is to make accurate forecasts for the Cyprus tourism demand. The necessity of these forecasts and the importance of tourism sector in Cyprus economic growth are in the front line of this work. These predictions can help the government and stakeholders to improve the economy and further develop this sector. In this empirical study, several models, already proposed by previously published articles on forecasting tourism demand, are applied. The reliability of the models is determined by three widely used forecast accuracy measures. Weighted forecast combinations are also employed in order to improve forecasting performance. Interesting outcomes and forecasts up to the end of 2018 are also presented.

Table of Contents

1. Introduction

Tourism has an enormous contribution to economy worldwide and without a doubt is one of the major economic sectors in Cyprus. Cyprus' sunny days that cover the largest part of the year, along with the excellent beaches' quality, constitute the traditional sun-and-sea model of this island. In 2016, the total contribution of Travel & Tourism to GDP and employment was 21.4% and 22.0%, respectively. Keeping the tourism in increasing rates implies keeping the economic growth in positive rates.

Motivated by these facts, one can think that variability and uncertainty in tourism can significantly affect the overall economy. Tourism is largely connected with government, through tax revenues, as well as households, through job opportunities and income, and other businesses, through sales and profits. Variability arrives from external circumstances, such as increasing terrorism in foreign countries, as well as climate change. In such periods when terrorism has increased substantially abroad, Cyprus is seen as 'safe-haven' destination, according to WTTC (2017). Nevertheless, uncertainty related with climate change and air quality may harm island's tourism. As pointed out in Rosselló-Nadal (2014), climate change, imposing high heat waves in summer, is bad news for warm destinations, especially when followed by dust, implying rather than optimal conditions and unpleasant experiences for tourists.

All the above evidence, highlights the necessity of the forecasts on tourism sector and the adoption of new policy measures in order to maintain Cyprus a desirable tourism destination. Hence, focusing on the forecasts, this study applies several models based on existing literature on tourism forecasts to stationary seasonally unadjusted monthly tourist arrivals in Cyprus from 2000 to 2017. The reliability of these models is judged by three of the mostly used measures of accuracy. Furthermore, in order to improve the estimations and make forecasts up to December of 2018, combinations of different forecasts are also implemented. Notable in this empirical analysis is the role of seasonality and stationarity.

The essay is organized as follows: Section 2 describes the models extensively used in literature. Section 3 refers to the seasonality and stationarity detection (Section 3.1) and provides the forecast combinations (Section 3.2) and the forecast accuracy measures (Section 3.3). Section 4 includes the empirical study: the data description (Section 4.1) and the results (Section 4.2) along with the forecasts for the upcoming period (Section 4.3). The last section concludes this essay.

2. Models

Many articles study the prediction of tourism demand in various countries. Seven of the most widely used model specifications are implemented in this paper. First of all, Seasonal AutoRegressive Integrated Moving Average (S*ARIMA*) model is the most common model used for forecasting tourist arrivals. The SARIMA model consists of non-seasonal and seasonal components and more importantly allows ARMA (Autoregressive Moving Average) to interact with seasonal effects. The second model employed is the Self-Exciting Threshold AutoRegression (*SETAR*) model which is an extension of autoregressive model that allows different behavior of time series in different regimes. Moreover, a model commonly used for time series data analysis is the Exponential Smoothing (*ES*) model which has the advantage of allowing exponentially decreasing weights over time. Relevant with ES, is the Holt-Winter exponential smoothing model (*HW*) which is basically an ES model that also takes into account three aspects: a level, a trend and a seasonal component. In addition, a notable model for this study is the Unobserved Components Model (*UCM*) that performs a time series decomposition into components such as trend, seasonal and cycle. Last but not least, two simple models are used in order to compare their efficiency with the aforementioned

ones: the linear regression (*LR*) model with monthly seasonal dummies as the predictor variables and the seasonal Naïve (*SN*) model that accounts for seasonality by setting each prediction to be equal to the last observed value of the same season.

The models used in the current analysis are proposed by the articles focusing on forecasting tourist arrivals. Claveria and Torra (2014) evaluate the forecasting performance of different models on overnight stays and tourist arrivals in Catalonia, amongst them, the SETAR model. Liang (2014) employs the SARIMA–GARCH model to analyze and predict tourism demand in Taiwan and makes a comparison with other models, such as ES and HW. In addition, Cho (2001), Hu et al. (2004) as well as Law (2001) implement, among other models, the ES. Saayman and Botha (2015) model tourism demand (tourist arrivals) using non-linear univariate time-series methods, such as UCM, also called the basic structural model (BSM), and compare them to the benchmark SN model and the popular SARIMA model. BSM model is also applied by Du Preez and Witt (2003) and Turnen and Witt (2001) for time-series forecasting. Moreover, SN model can be found in Hu et al. (2004), Law (2004), Petropoulos et al. (2003) and Song et al. (2000) which are related with econometric modeling and forecasting. Last but not least, Mamula (2015) examines the forecasting tourism demand in Croatia by incorporating seasonal dummies in LR model to capture the seasonality effect.

3. Methodology

3.1 Seasonality and Stationarity

Seasonality is a key feature of many tourist arrival series. Several factors, such as weather and holiday days, have an impact on the arrivals of tourists. In addition to this, time series suffer from non-stationarity. The models implemented require stationary series. To investigate these issues, the autocorrelation function (ACF) is employed and plots of the series against their ACF are created and presented in Figure 1. ACF measures the relation between current and past series values.

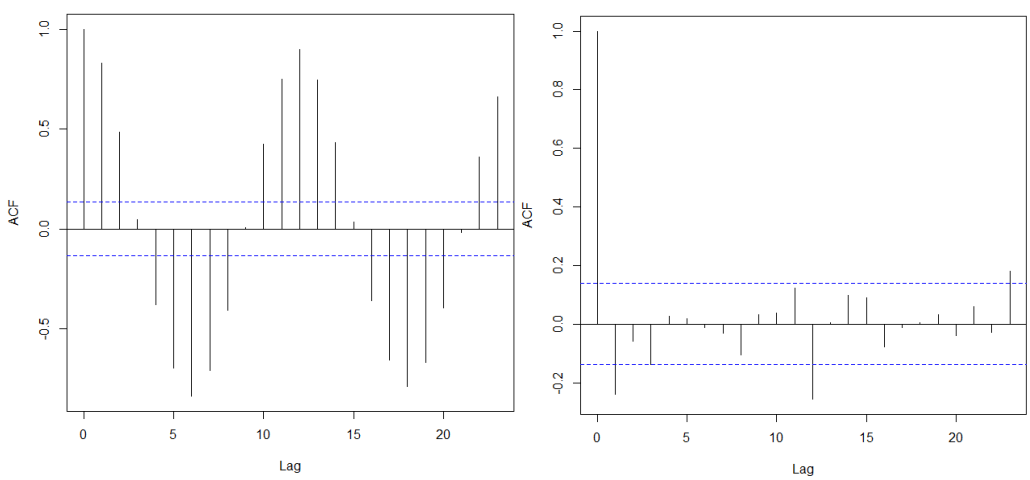

Figure 1. ACFs of original (non-stationary) (left part) and stationary (right part) time series.

The pattern shown in the left part of Figure 1 ensures that there exists strong seasonality in these time series, mainly due to the sun-and-sea Cyprus product which strongly enhances summer tourism. Following Liang (2014), the first order difference and the first seasonal difference of the logarithms of the time series data are taken to attain stationarity of the tourist arrival series. Turning to the stationary time series in

the right part of Figure 1, it is obvious that the ACFs are reduced. Seasonality is embodied within the models the way described in the previous section.

3.2 Forecast Combinations

In order to improve the forecasting reliability, combinations of the seven models are considered. Forecast combinations are developed in Stock and Watson (2004), Timmermann (2005) as well as Smith and Wallis (2009). The mostly used forecast combination is the weighted average of forecasts. Firstly, the equally-weighted or *simple* average is implemented where all the models' forecasts (f_i) have the same weight. The combined forecast (f^c) is given by

$$f^c = \frac{1}{8}\sum_{i=1}^{8} f_i \ .$$

Secondly, the *variance-based* or Inverse Mean Squared Error (Inverse MSE) weighted average is applied where the models have different weights according to their accuracy. The combined forecast is given by

$$f^c = \sum_{i=1}^{8} \frac{\frac{1}{MSE_i}}{\sum_{i=1}^{8} \frac{1}{MSE_i}} f_i \ .$$

3.3 Forecasting Accuracy

Forecasts ($\widehat{y_t}$) are estimated based on the training set and then are compared with the testing set (y_t) in order to extract the forecasting errors. The performances of models based on these errors are judged by three mostly used accuracy measures: the mean absolute percent error (MAPE), the root mean square percentage error (RMSPE) and the mean absolute deviation (MAD), given by

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{y_t})^2}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|y_t - \widehat{y_t}|}{|y_t|}$$

$$MAD = \frac{1}{T}\sum_{t=1}^{T}|y_t - \widehat{y_t}| \ .$$

4. Empirical Study

4.1. Data

This study implements the seven models, described in Section 2, on Cyprus tourism data, focusing on tourist arrivals of non-residents. Tourist arrivals consist the most popular time series related to forecasting tourism demand in literature. The data, presented in Figure 2, are extracted from Eurostat Database for the period of January 2000 to October 2017, consisting of 241 observations.

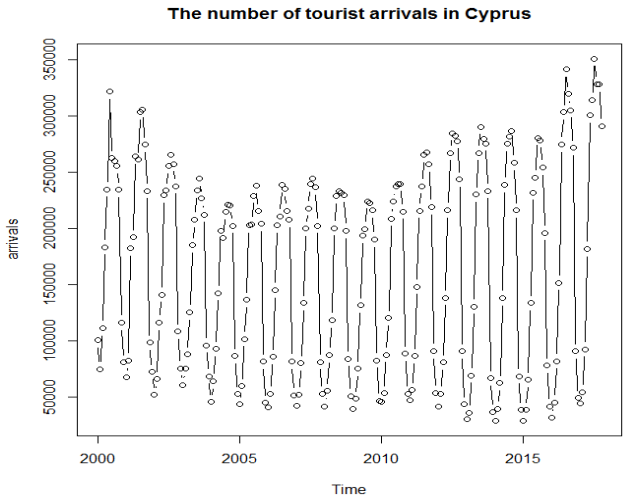

Figure 2. Time series plot of the number of tourists' arrivals in Cyprus.

Using the monthly frequency provides a large time-series sample which is essential for reliable estimates of the unknown models' parameters and small forecasting errors compared to sample of quarterly or annual frequencies. Furthermore, monthly frequency provides useful information related to forecasts that are connected with different months of the year. Hence, from the economic point of view, as emphasized in Song and Li (2008), economic policy makers desire the prediction of tourism demand in monthly basis in order to think about e.g. how to stimulate winter tourism so as to cover fixed costs related to this sector. Similarly, it can also help tourist agents in order to decide which tourist units to stay open in which months of the year and manage resources, such as staffing ad stock arrangement.

In Table 1, the summary statistics for the tourist arrivals series are obtained. Data from January of 2000 to December of 2014 compose the training set, while the remaining data constitute the testing set. The models are estimated based on the training set and then forecasts are computed up to 36 steps ahead, corresponding to the out-of-sample period of January 2015 to December 2017.[1]

Series	**Tourist arrivals**
Sample	From January 2000 to October 2017
Observations	214
Mean	163,675
Median	191,132
Maximum	351,249
Minimum	28,883
Std. Dev.	90,975.94
Skewness	0.03
Kurtosis	1.58
Jarque-Bera	17.96

Table 1. Summary statistics of tourist arrivals.

[1] The results shown in the following subsections are robust to a shorter out-of-sample period given by January 2016 to December 2017.

4.2. Results

Table 2 presents the results of the accuracy measures applied on the forecasts of the models, using R.[2] RMSE and MAD select SARIMA model as the most suitable for tourism analysis, while MAPE prefers the ES model. As a result, SARIMA model provides the most accurate forecasts and outperforms the others according to two out of the three criteria.

	RMSE	MAPE	MAD
SARIMA	**0.077**	1.288	**0.060**
SETAR	0.088	1.447	0.067
ES	0.086	**1.069**	0.067
HW	0.112	4.997	0.090
UCM	0.093	1.894	0.068
LR	0.091	1.352	0.069
SN	0.144	5.636	0.116

Table 2. Accuracy measures' results of the forecasting models. Values in bold indicate the models that minimize each measure.

However, since ES also minimizes the MAPE criterion, it is recommended to use forecast combinations. Simple average gives equal weight to the seven models while the variance-based average gives the weights shown in Table 3. As expected SARIMA model has the highest value of weight.

	SARIMA	SETAR	ES	HW	UCM	LR	SN
First	0.209	0.164	0.169	0.100	0.146	0.151	0.061

Table 3. Weights of variance-based forecast combinations.

Then, in Table 4, the three criteria are applied so as to check the forecast combinations' accuracy compared with the forecasts of the simple average and the best model (SARIMA). On the one hand, RMSE and MAD measures agree that the SARIMA model is the most successful one, while on the other hand, MAPE chooses variance-based forecast combination.

	RMSE	MAPE	MAD
Simple Average	0.090	1.402	0.070
Variance-based Average	0.086	**1.121**	0.065
Best (SARIMA)	**0.077**	1.288	**0.060**

Table 4. Accuracy measures' results of the forecast combinations and SARIMA model. Values in bold indicate the forecasts that minimize each measure.

Focusing on Figure 3, where the real data of the testing set along with the different forecasts are presented, it is clear that all the forecasts are close to the observations and not many differences exist between the variance-based forecast combination and the SARIMA estimates. There is a discrepancy

[2] For the SARIMA model, in the training set a SARIMA(5,0,2)x(2,0,1) is applied. These numbers of order are selected according to AIC criterion. For the SETAR model, values 3 and 4 are chosen as the orders of AR terms.

between forecasts and actual values but one can see from Figure 3 that in many periods the forecasts coincide with actual values e.g. in some periods they might be lower and in other periods they might be higher.

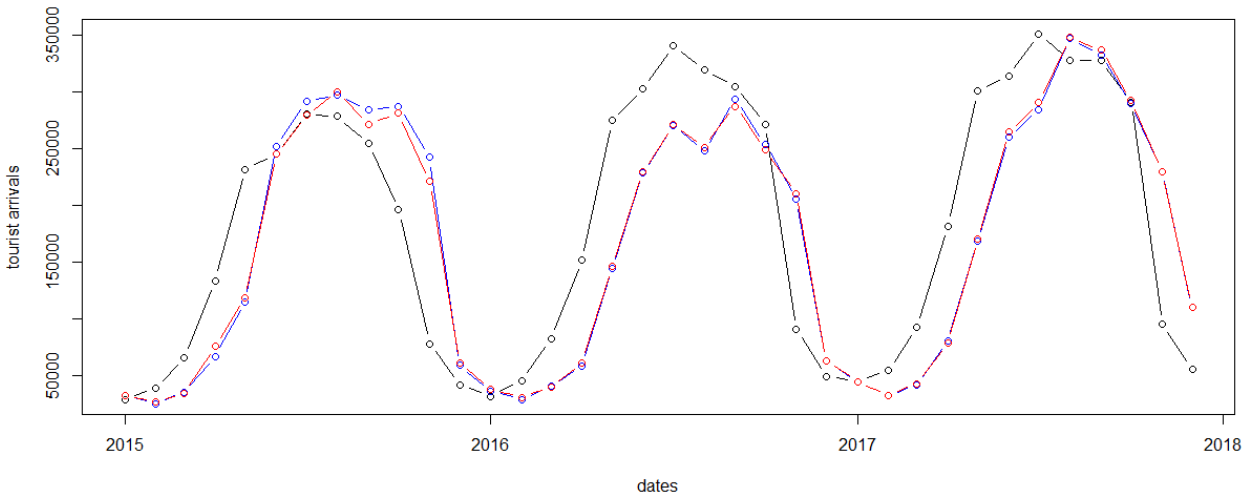

Figure 3. Black line represents the real data on tourist arrivals. Blue line represents the forecasts of SARIMA model and red line represents the forecasts from the variance-based combination of all models.

Comparing the yearly aggregated forecasts with yearly aggregated monthly actual tourist arrivals in Table 5, it becomes clear that estimates of forecast combination are closer to the real ones vis-à-vis forecasts obtained by SARIMA model. Driven by all the above results, forecasts extracted from the variance-based combination, with the highest value of weight on SARIMA model, are considered as the most accurate ones.

	SARIMA	**Variance-based Combination**	**Actual**
2015	1,990,263	1,948,811	1,874,322
2016	1,872,695	1,878,203	2,268,176
2017	2,223,140	2,242,662	2,438,932

Table 5. Yearly aggregated forecasts vis-à-vis yearly aggregated monthly actual tourist arrivals

4.3. Forecasts

Using the variance-based forecast combination, the forecasts for tourist arrivals from January of 2018 until December of 2018, which embody seasonality, are found in Table 5.[3] These forecasts lead to notable outcomes. In the whole dataset the maximum number of tourist arrivals is close to 351 thousands while the minimum is almost 29 thousands, as reported in Table 1. Observing Table 5, the forecasts of tourist arrivals on July, August and September 2018, around 392, 370 and 369 thousands, respectively, overcome this maximum number and the minimum forecasting value of almost 51 thousands is far above the minimum existing value. These outcomes provide evidence for increasing rates in tourist arrivals.

[3] For the full sample a SARIMA(3,0,3)x(2,0,2) model is applied.

Month	Forecasts on tourist arrivals
2018M01	50,915
2018M02	62,682
2018M03	102,193
2018M04	202,027
2018M05	336,224
2018M06	348,200
2018M07	392,172
2018M08	370,225
2018M09	368,625
2018M10	330,855
2018M11	106,437
2018M12	62,939

Table 5. SARIMA forecasts for tourist arrivals from November 2017 to December 2018

5. Conclusion

In this study, seven different models were examined on tourism demand forecasting. Based on widely used forecast accuracy measures, the forecasts extracted from the variance-based forecast combination outperform the others. Therefore, this combination is implemented to make forecasts up to December 2018 and the results indicate that tourist arrivals will keep having positive growth rates.

At this point, it is important to mention why these forecasts are pivotal. First of all, findings of this paper might prove to be useful to governmental services, stakeholders of Cyprus Tourism Organization and policy makers while their purpose is to distribute effectively the existing resources and also help the potential investors to plan their projects with greater certainty. Secondly, forecasting decreased numbers of tourists for some months, can assist in adopting relative actions to attract more visitors. January is traditionally the month with the fewest number of tourist arrivals while July brings the highest number of tourists. In this context, a better seasonal distribution of tourism activity is of great importance. As mentioned before, excessive temperatures and air quality have to be considered as threats in order not to lean exclusively on summer tourism but to support and enforce tourism during other periods.

Cyprus has been significantly developed in Services sector in all regions. Newly established and one-of-a-kind restaurants, bars, shops, relaxing places such as spa and wellness retreats can bring tourism throughout the year in different areas if promotions and advertisements take place in the right way. So, what is actually needed, as highlighted in Clerides and Pashourtidou (2007), is a co-ordination of all bodies involved in tourism. A synergy of government and private sector is required with costly but valuable investments, not only focusing on public transport but also, offering attractive packages inviting tourists to get acquainted to another side of Cyprus. Taking advantage of island's favorable winter, weather conditions and creating a plan for tourism in a 12-month basis can unfold a whole new vision for Cyprus tourism.

Last but not least, relative to tourism's economic contribution, this sector is an insurance policy in case of bad times. Specifically, under economic recessions additional revenues coming in through tourism can support the island's industries. That's a phenomenon we observe in Cyprus, where tourism helps the

recovery of Cyprus economy after the financial crisis of 2013. Driven by all the above considerations, accurate prediction of the tourism demand is vital to the tourism industry, notably when tourism is the "workhorse" for one country's economic growth, as in the case of Cyprus.

References

Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs time series models, In Economic Modelling, 36, 220-228.

Clerides, S. & Pashourtidou, N. (2007). Tourism in Cyprus: Recent Trends and Lessons from the Tourist Satisfaction Survey, Cyprus Economic Policy Review, 1, issue 2, 51-72.

Cho, V. (2001). Tourism forecasting and its relationship with leading economic indicators. Journal of Hospitality and Tourism Research, 25,399–420.

Du Preez, J., & Witt, S. F. (2003). Univariate versus multivariate time series forecasting: An application to international tourism demand. International Journal of Forecasting, 19, 435–451.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages, ONR Research Memorandum, Carnegie Institute of Technology 52.

Hu, C., Chen, M., & McCain, S. C. (2004). Forecasting in short-term planning and management for a Casino Buffet Restaurant. Journal of Travel & Tourism Marketing, 16, 79–98.

Law, R. (2001). The impact of the Asian financial crisis on Japanese demand for travel to Hong Kong: A study of various forecasting techniques. Journal of Travel & Tourism Marketing, 10, 47–66.

Law, R. (2004). Initially testing an improved extrapolative hotel room occupancy rate forecasting technique. Journal of Travel & Tourism Marketing, 16, 71–77.

Liang, Y. (2014). Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan, In Computers & Industrial Engineering, 74, 111-119.

Mamula, M. (2015). Modelling and Forecasting International Tourism Demand – Evaluation of Forecasting Performance, International Journal of Business Administration, 6.

Petropoulos, C., Patelis, A., Metaxiotis, K., Nikolopoulos, K., & Assimakopoulos, V. (2003). SFTIS: A decision support system for tourism demand analysis and forecasting. The Journal of Computer Information Systems, 44, 21–32.

Rosselló-Nadal, J. (2014). How to evaluate the effects of climate change on tourism, Tourism Management, 42, 334-340.

Saayman, A., & Botha, I. (2015). Non-linear models for tourism demand forecasting, Tourism Economics, 23, Issue 3, 594 – 613.

Smith, J., & Wallis, K. F. (2009). A Simple Explanation of the Forecast Combination Puzzle, Oxford Bulletin of Economics and Statistics, 71, 331–355.

Song, H., & Li, G. (2008). Tourism Demand Modelling and Forecasting: A Review of Recent Research Tourism Management, 29 (2), 203-220.

Song, H., Romilly, P., & Liu, X. (2000). An empirical study of outbound tourism demand in the UK. Applied Economics, 32, 611–624.

Stock, James H., & Watson, Mark W. (2004). Combination forecasts of output growth in a seven-country data set, Journal of Forecasting, 23, 6, 405—430.

Timmermann, Allan G. (2005). Forecast Combinations, CEPR Discussion Paper No. 5361.

Turner, L. W., & Witt, S. F. (2001). Forecasting tourism using univariate and multivariate structural time-series models. Tourism Economics, 7, 135–147.

WTTC (World Travel & Tourism Council) (2017) Travel & Tourism Global Economic Impact & Issues 2017.