33rd European Modeling & Simulation Symposium
18th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0029 ISBN 978-88-85741-57-7 © 2021 The Authors.
doi: 10.46354/i3m.2021.emss.006

Forecasting Tourism Demand in Greece Using Time Series Forecasting

Eleni Saltsidou[1,*], Maria Drakaki[2]

[1]International Hellenic University, 14th km Thessaloniki -N.Moudania, Thessaloniki, GR-57001, Hellas (GR)
[2]International Hellenic University, 14th km Thessaloniki -N.Moudania, Thessaloniki, GR-57001, Hellas (GR)
[*]Corresponding author. Email address: esaltsidou@ihu.edu.gr

Abstract

Tourism is one of the most important industries of the Greek economy and is a key to economic growth. In 2019, 27.53 million tourists arrived in hotel and camping accommodations of the country. While these arrivals help the economic growth of the country, there is a need to meet the expectations of tourists in the context of the provided services. The tourism industry should make investments in the infrastructure of the destinations in order to accommodate and manage the tourist flows. These investments include transportation services, logistics, accommodation and health services. Therefore, different methods have used tourism industry data to come up with accurate forecasts of tourist arrivals. The research on tourism demand supports decision making bodies to develop and improve practices that contribute to tourism development. The main aim of the paper is to provide time series forecasting models of tourist arrivals in Greece in order to assist decision making bodies and stakeholders to develop and improve practices that contribute to tourism development. Moreover, this paper aims to contribute to the research field of tourism demand forecasting with a case study that concerns Greece. Historical data of tourist arrivals to the Ionian Islands, obtained from ELSTAT for the time period 2010-2018, have been used in the time series forecasting models used in this research.

Keywords: Tourism demand; forecasting; time series; Greece; tourism services;

1. Introduction

Tourist arrivals are an essential factor to understand the tourism demand industry and its trends. Accurate forecasts of tourism demand are necessary for stakeholders so that they can organize an efficient plan to properly distribute their resources. Tourism demand forecasting has attracted research attention in the recent years. Many methods have been developed in order to produce an optimal model that has an accurate forecasting performance. Most applications are concerned with time series models that aimed to forecast the tourist flows. The models that are mostly utilized are ARIMA and SARIMA, yet many applications and experiments were developed with other important time series models such as the exponential smoothing and the Naïve method. Besides, tourism forecasting has been implemented with econometric models which are highly accurate. Also, in the past decades, many applications involved Artificial Intelligence (AI) methods. Especially, Neural Networks (NN) and Support Vector Machine (SVM) have been widely used for tourism demand forecasting (Xu et al., 2016).

Many studies have been conducted involving tourism in Greece. There are two research approaches used in the context of the tourism industry. One approach concentrates on tourism patterns and impacts of

© 2021 The Authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

tourism development, demonstrating the reshaping of the local socio-economic system due to the tourism market. The other approach concentrates on the attributes that influence the tourism industry and the attributes that are influenced by tourism such as infrastructure services, tourist agencies, and entrepreneurs (Galani-Moutafi, 2004).

In the 90s and 00s, the scope of research was to propose changes that were about to emphasize the natural, cultural, and historical resources or even the development of alternative types of tourism along with the application of forecasting tourism demand methods (Galani-Moutafi, 2004). These studies have focused mainly on islands or areas with summer tourism (Koutras et al., 2016).

This study aims to highlight the importance of forecasting and the accurate evaluation of the results in order to create an efficient plan for the distribution of tourism flows in Greece. In addition, this paper attempts to contribute to the forecasting field predicting the tourist arrivals to the islands that are in great demand which has not been done by other authors and the forecasts involve periods with complex problems such as migration flows, economic crisis and pandemics. From a modelling scope, this study attempts to provide a model that can generate efficient forecasts. The predictions of this research are on a monthly basis. It can be useful for the development of short-term planning for the stakeholders and those who work in the tourism sectors in order to better organize their services. In addition, more accurate and consistent results can be achieved with monthly data since it consists of a large number of observations, and the fluctuations that concern seasonality are more easily observable. Lastly, the research can be generalized for other countries including EU countries since the models generate accurate results.

2. Literature review

Tourism demand forecasting involves quantitative and qualitative approaches. Qualitative approaches are based on information derived from human opinions and experiences (Xu et al., 2016; Abellana et al., 2020). These approaches are not efficient for generalization, therefore, are not commonly utilized for research purposes. Yet, they are very efficient in some cases which involve the understanding of the connections between the demand and the factors which make tourists choose a certain destination (Xu et al., 2016; Abellana et al., 2020). On the other hand, quantitative tourism demand forecasting models are based on mathematical approaches involving numeric data. These models have the scope to predict future values that depend on past performances (Abellana et Al., 2020). Also, forecasting models have focused on both long-term and short-term forecasts. Short-term forecasts are daily predictions. The decision regarding which is the best forecasting method depends on the specific case (Bi et al., 2020).

There is no single model that outperforms other models for all problems (Claveria et al., 2013). Instead, the optimal model depends on the particular problem to be solved and the available data.

The quantitative forecasting methods can be categorized into three main categories: Time series models, AI models, and econometric approaches (Peng et al., 2014).

2.1. Time series models

Time series models are based on historical data (Song, 2019). They are a sequence of data points, divided in time intervals. The basic concept of time series analysis involves the past state of patterns which will continue into the future (Rathnayaka et al., 2016). Through these models, scientists attempt to recognize the trends, slopes, and the seasonality of a time series (Song, 2019). When a specific pattern occurs in a time series, time series models can generate future values (Song, 2019; Álvarez-Díaz et al., 2019). Besides, time-series models are widely used by the researchers because they require less time and less cost to come up with accurate results (Álvarez-Díaz et al., 2019).

Time series forecasting can be utilized in many fields such as pattern recognition, engineering, finance, etc. As a result of the applicability and the highly volatile behavior of the data in the past decades, many studies have focused on improving the time series forecasting models or combining many models to produce a hybrid optimal model (Rathnayaka et al., 2016).

Time series models can be divided into two categories: the basic models and the advanced models (Song, 2019). The basic models include Naïve methods and exponential smoothing. These models are easy to implement and have a great ability to display patterns through data points (Song, 2019). Also, the advanced models include ARIMA and SARIMA which have been widely used in many applications. Time series models have been extensively used in tourism demand forecasting (Song, 2019; Petrevska, 2017; Claveria & Torra, 2014). Yet, the most dominant method is the ARIMA (or SARIMA) concerning the tourism demand forecasting (Chu, 2008; Du Preez & Witt, 2003; Claveria & Torra, 2014; Song, 2019).

2.2. Evaluation metrics

The performance of every forecasting model needs to be measured and evaluated (Goh & Law, 2002). There are many methods and metrics which are utilized to measure the accuracy and to interpret the error (Goh & Law, 2002). The mean absolute percentage error (MAPE), the mean-squared error (MSE), and the root-mean-squared error (RMSE) have been extensively utilized in this literature. The key to an efficient evaluation of a model is the minimization of the error (Goh & Law, 2002).

The equation for the MAPE is displayed below.

$$\frac{\sum\left|(A_t - F_t)/A_t\right|}{n}. \qquad (1)$$

Where At is the actual value of period t, Ft is the forecasted value in period t and n is the number of periods.

3. Methodology

The tourism sector plays a highly important role for Greece. The main objective of this research is to make long-term estimations about tourism demand in Greece by introducing several time-series models. With this objective, we aim to improve the forecasting of tourism demand in Greece and to contribute to the research in this scientific field in order to support the tourism and the associated industries.

The main aim of the research is to provide quantitative results in terms of forecasts about the trends for tourism demand in Greece during the summer period and to determine the peak periods. Moreover, the paper aims to find which time-series model serves the best in every case. The predictions of this research are on a monthly basis. It is more interesting to predict monthly tourism demand data rather than yearly. Also, it can be useful for the development of short-term planning for the stakeholders and those who work in the tourism sectors in order to better organize their services.

For the scope of this research, the data that has been utilized includes tourist arrivals. The datasets include data on monthly tourist arrivals of natives and foreigners for the period between January 2010 and December 2018. The datasets concern the prefecture of the Ionian Islands that is a summer destination. The datasets are collected from ELSTAT, the Hellenic Statistical Authority, which is an authority for sharing statistical inferences and data observing several rules and patterns.

The R studio was selected as the implementation tool, because it provides a lot of functionality useful for time-series analysis. The general layout of the methodology consists of two steps. The first step involves the data cleaning. The second step is the implementation of every implemented model on the dataset. This research has focused on forecasting monthly tourist arrivals. The experiments aimed to train the models with data from January 2010 to December 2018 and then forecast the monthly arrivals of 2019. The methods that have been utilized in this research are presented below.

3.1. Naïve/Seasonal Naïve

For naïve forecasts, the procedure is to set all the forecasted values to be one of the last observations. For seasonal naïve, the procedure is the same as above, with the difference that the last observed value from the same season of the year is set as the forecasted value. Naïve 1 is:

$$F_t = x_{t-1}$$

Where F_t is the forecasted value at time t and x_t is the observed value at time t.

3.2. ARIMA models

ARIMA is one of the widely used methods in the tourism industry and is a method that usually generates accurate results. In this method, a series is changed to covariance stationary conditions and after that, it is identified, assessed, analyzed and forecasted. The ARIMA is written as ARIMA (p,q,d) where p expresses autoregression, d differencing and q is for moving average. The model that has been utilized in this research is SARIMA (seasonal ARIMA). It has the form SARIMA (p,q,d)(P,Q,D) where the capitalized letters stand for the seasonal components.

The moving average models average the data adjusting some weights to each value of N values (Goh, 2011). The value of N is based on a certain series and its content. For example, N=12 when the series contains annual data (Goh, 2011).

The original dataset for tourist arrivals at the Ionian Islands is shown in Figure 1. Besides the seasonal component, a large drop can be observed in 2015. In 2015 in Greece there was a situation concerning limitations on banking systems. Also, another factor that influenced tourist arrivals was the migration flows. In 2015 in Greece, massive arrivals of refugee inflows were observed (Kasimati, 2016).

These inflows could potentially have impacts on tourist arrivals since many infrastructures that accommodate refugees are based on islands such as Chios or Lesvos. Assuming that those situations had impacts on tourism demand in 2015, the decrease in tourist arrivals which is displayed in Figure 1 could be explained.

From the adf.test function which stands for augmented Dickey–Fuller test in order to check the stationarity of the time-series a p-value is generated for the Ionian dataset. The p-value is 0.01<0.05 which means that the series is stationary. The year 2015 was considered an outlier. The transformed time series after the data cleaning is shown in Figure 2.

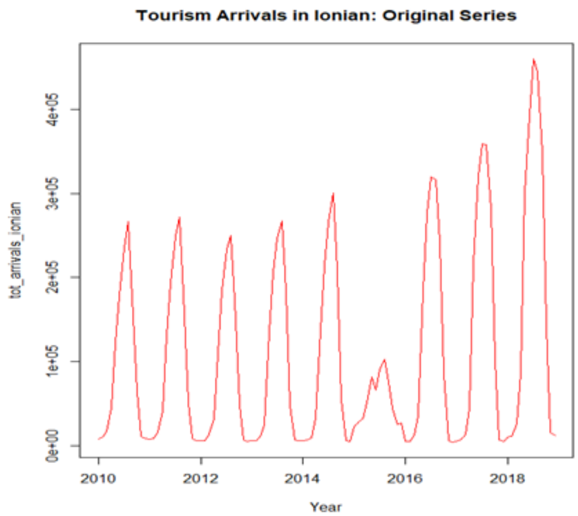

Figure 1. Tourism arrivals in the Ionian Islands (Original Series)

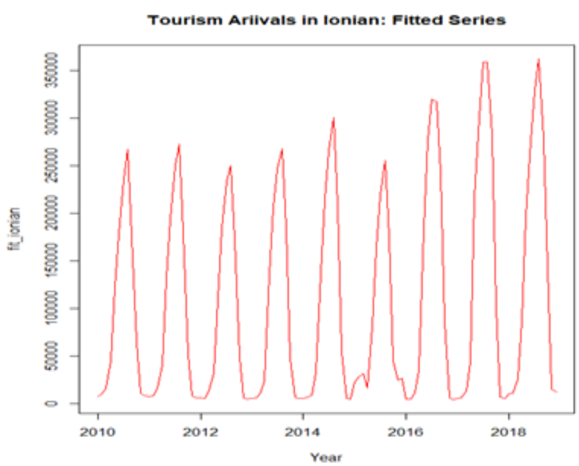

Figure 2. Tourism arrivals in the Ionian Islands (Fitted series)

Some limitations were encountered. As it was observed, 2019 was the year with the most tourist arrivals from all the periods for which data was collected. Also, there is an increase through the years up to 2019, with 2019 to be the most profitable year in terms of arrivals. This increase was not recorded through the training phase so it is assumed that there will be a difference in forecasted values and the actual values since the values of 2019 were never observed from the models.

Another limitation is the difficulty to make forecasts concerning the year 2020. Coronavirus has impacted millions of lives on earth (Sigala, 2020). One of the many impacts is that Greece has very small rates on tourism arrivals this year due to Covid-19 and pandemic situation (OECD, 2020). Involving 2020 to the forecasting horizon could bring low MAPE rates since the actual value of 2020 would be much lower than the forecasted value given the fact that the model had not been trained with such values. In other words, we can call 2020 an "outlier" year and especially for forecasting problems, so it was decided not to be utilized as a test set, since this could yield small accuracy rates.

4. Results

The experiments involve 12-month predictions. The 12-month forecasts concern the year 2019.

The seasonal Naïve, and SARIMA for different values of (p,q,d)(P,Q,D) were utilized at the dataset. 108 data points were utilized for training purposes which were the monthly arrivals between 2010 and 2018. The next 12 data points were used as a test set. The decision of the best model that serves every case is based on the evaluation criteria and the accuracy metrics. Firstly, the model selection for every pair of parameters at ARIMA model was chosen based on the smaller MAPE. The optimal models are the seasonal ARIMA $(1,0,3)x(0,1,1)_{12}$ and the Naïve method with the smallest MAPE. The 12-month predictions for every model are shown in Figures 3, 4 and 5, respectively and the figures display the existing time series along with the predicted year in blue colour. Previous studies of forecasting are synthesized in order to rank the forecasting models to their average MAPE. The seasonal Naïve method gave MAPE equal to ~19% which is an acceptable error rate. This error rate shows that the forecasted values are close enough to the actual values which means that both models performed well in this case. This could also be seen from the figures since the forecasted curve is quite close to the previous observations. In addition, ARIMA $(1,0,3)x(0,1,1)_{12}$ and ARIMA$(1,0,2)x(0,1,1)_{12}$ gave MAPE 22.51 and 23.32, that shows that the SARIMA model performs well with short-term forecasts. Yet, the seasonal naïve model outperformed the SARIMA model.

5. Conclusions

The main aim of this research was the forecasting of tourist arrivals in Greece and the assessment of the accuracy of the implemented models' performance, as well as whether there is a model that outperforms all others. Tourist arrivals in the Ionian Islands were utilized for testing in order to find the most accurate model in terms of tourism demand forecasting. We performed an analysis on monthly arrivals covering the period from January 2010 to December 2019. With the utilization of these models, it is better to keep our predictions at short forecasting horizons. After testing different tests each time and depending on the available datasets, the aim is to find the optimal model for the specific case at hand. The findings of this study have proven to be a good step towards improving the forecasting accuracy as far as the tourist destinations in Greece are concerned. The application of the proposed models or even a combination of the models to a wider range of tourist destinations is worth further investigation. Additionally, future research could test a combination of AI models with time-series models in order to produce a hybrid model that behaves well with datasets that include unexpected events such as the economic crisis in a country, pandemics such as the Covid-19 situation and other factors such as migration flows to touristic destinations.

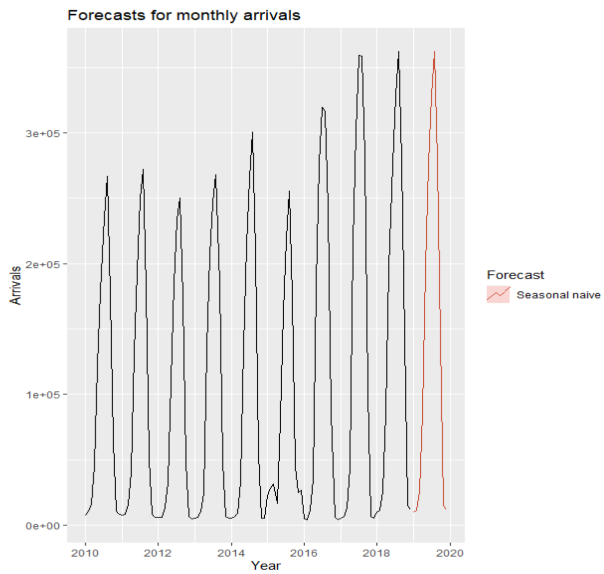

Figure 3. Forecasts for tourist arrivals in the Ionian Islands (seasonal Naive model)

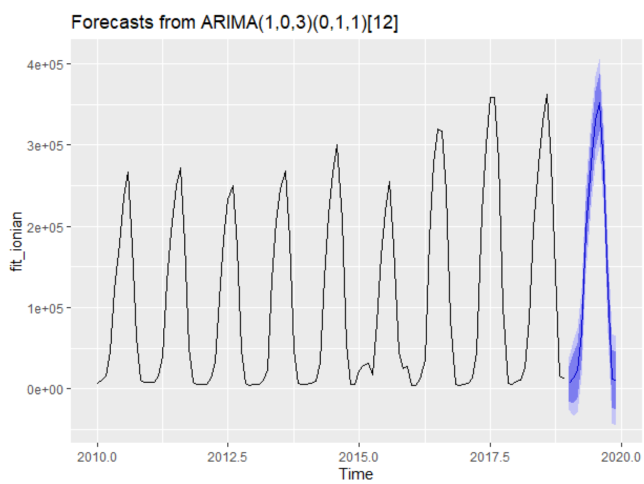

Figure 4. Forecasts for tourist arrivals in Ionian Islands using the ARIMA (1, 0, 3) x (0, 1, 1)$_{12}$

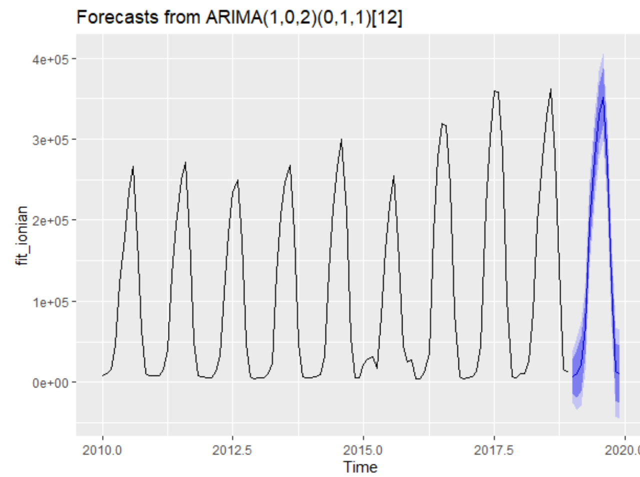

Figure 5. Forecasts for tourist arrivals in Ionian Islands using the ARIMA (1, 0, 2) x (0, 1, 1)$_{12}$

The research can be generalized for other countries including EU countries since the models generate accurate results. The research methodology could be applied to forecast tourism demand on European or international destinations where the datasets of certain destinations present the same patterns as the data used to train the models developed and used in this paper.

References

Abellana, D. P. M., Rivero, D. M. C., Aparente, M. E., & Rivero, A. (2020). Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario. *Journal of Tourism Futures.*

Alikhani R., Torab S.A., Altay N. (2019). Strategic supplier selection under sustainability and risk criteria. *International Journal of Production Economics*, 208, 69–82.

Álvarez-Díaz, M., González-Gómez, M., & Otero-Giráldez, M. S. (2019). Forecasting international tourism demand using a non-linear autoregressive neural network and genetic programming. Forecasting, 1(1), 90-106.

Bi, J. W., Liu, Y., & Li, H. (2020). Daily tourism volume forecasting for tourist attractions. Annals *of Tourism Research*, *83*, 102923.

Chu, F. L. (2008). Analyzing and forecasting tourism demand with ARAR algorithm. *Tourism Management*, *29*(6), 1185-1196.

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

Du Preez, J., & Witt, S. F. (2003). Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, *19*(3), 435-451.

ELSTAT (https://www.statistics.gr/)

Goh, C., & Law, R. (2011). The methodological progress of tourism demand forecasting: a review of related literature. *Journal of Travel & Tourism Marketing*, *28*(3), 296-317.

Kasimati, E., & Ioakeimidisb, P. (2019). Accessible tourism in Greece: What is the current status?. ASSOCIATE EDITOR, 1, 33-42.

Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, *45*, 181-193.

Sigala, M. (2020). Tourism and COVID-19: impacts and implications for advancing and resetting industry and research. *Journal of Business Research*.

Song, H., Li, G., Witt, S. F., & Athanasopoulos, G. (2011). Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27(3), 855-869.

Xu, X., Law, R., Chen, W., & Tang, L. (2016). Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs. *CAAI Transactions on Intelligence Technology*, *1*(1), 30-42.