Technical Assistance action to support tourism planning and policy for the promotion of sustainable tourism development in Greece

**Component I: Tools for policy making**

Activity I.1.2: Develop monitoring and reporting tools for Greek tourism: a system documenting and forecasting tourism development

**Forecasting Tourism Development**

**Final Report**

**Table of Contents**

[Executive Summary 3](#_Toc491819277)

[1. Introduction 5](#_Toc491819278)

[2. The data set. Empirical Analysis and Results 8](#_Toc491819279)

[2.1. Data collection, data base modelling 8](#_Toc491819280)

[2.2. Descriptive Statistics 11](#_Toc491819281)

[2.3 Analysis by month 11](#_Toc491819282)

[2.4. Analysis by Seasons 14](#_Toc491819283)

[2.4.1. Definition of Seasons 15](#_Toc491819284)

[2.4.2. Analysis of Seasons 16](#_Toc491819285)

[2.5. Correlations 22](#_Toc491819286)

[2.6. Analysis of Arrivals by country of origin 25](#_Toc491819287)

[2.6.1. The data set 25](#_Toc491819288)

[2.6.2. Most significant countries 27](#_Toc491819289)

[2.6.3. Presentation of Time-series and Trend 33](#_Toc491819290)

[Countries with almost zero trend 34](#_Toc491819291)

[Countries with positive trend 38](#_Toc491819292)

[Country with negative trend 43](#_Toc491819293)

[3. Forecasting 44](#_Toc491819294)

[3.1. Introduction 44](#_Toc491819295)

[3.2 Short review on forecasting methods in tourism 46](#_Toc491819296)

[3.3. Methods 48](#_Toc491819297)

[3.3.1. Exponential Smoothing 48](#_Toc491819298)

[3.3.2. The Box-Jenkins approach (ARMA) – SARIMA framework 48](#_Toc491819299)

[3.4. Model fitting 50](#_Toc491819300)

[3.5. Evaluation of forecasting accuracy 51](#_Toc491819301)

[3.6. Application and Results 53](#_Toc491819302)

[3.6.1. Tourist Arrivals 53](#_Toc491819303)

[3.6.2. Nights – Total 61](#_Toc491819304)

[3.6.3. Arrivals – Residents – Total 65](#_Toc491819305)

[3.6.4. Arrivals – Non-Residents – Total 67](#_Toc491819306)

[3.7. Conclusion 70](#_Toc491819307)

[4. Final remarks 71](#_Toc491819308)

[4.1. Summary 71](#_Toc491819309)

[4.2. Conclusions and next steps 72](#_Toc491819310)

[Appendices 76](#_Toc491819311)

[Appendix I. Descriptive Statistics of all variables included in the data set 76](#_Toc491819312)

[Appendix II. Plots of Nights-Total, Arrivals-Total by Month 85](#_Toc491819313)

[Appendix III. Mean and ANOVA tests for Seasons 109](#_Toc491819314)

[Appendix IV. Literature Review – Tourism Forecasting 111](#_Toc491819315)

[References 121](#_Toc491819316)

# **Executive Summary**

This report summarizes the main results of the Activity I.1.2: “Develop monitoring and reporting tools for Greek tourism: a system documenting and forecasting tourism development” of the Component 1 “Tools for strategic policy making” of the Technical Assistance Action. The Workplan stipulates as output of this Activity: “A concrete proposal on how to develop a system documenting and forecasting tourism development not only as a policy tool for the Ministry of Tourism but also as a campaigning tool for Greek National Tourism Organisation (GNTO)”. The main parts of the output of this Activity are (1) the report on the empirical analysis of tourism flow data and (2) the results from the investigation of forecasting methods applied to predict tourism flows including a proposal of a methodology which can be used routine-wise for forecasting the number of arrivals of visitors as well as the number of overnights.

Forecasts of tourism statistics are of crucial importance and interest for all institutions which are actors in the Greek tourism such as political authorities and tourism businesses. The interest in forecasting results concerns primary variables that describe the tourism flow like the number of visitors (arrivals, overnights), the duration of stay, and the expenditures of the visitors. Deeper analyses refer to details of the tourism flows like the inbound flows from certain countries of residence, flows according to purpose of travel, to regions of destination, and others. Deeper analyses may also refer to the generation of flows as a function of factors like disposable income of visitors, price levels in Greece and alternative destinations, travel costs, exchange rates, Greek marketing expenditure, events like political unrests, and others.

Forecasting methods for the extrapolation of time series like the monthly observations of tourism flows are based on rather simple models like exponential smoothing and time-series models like AR, MA, ARMA, and ARIMA; the forecasts are typically only function of the past of the time-series. Forecasting methods which allow to make use of interactions between the variable to be predicted and factors which determine the data-generating process are more demanding both with respect to the data base and the models to be used such as Vector Autoregressive models, Error Correction models, Time Varying Parameter (TVP) models, etc.

It was decided that in a first stage of the project, the focus should be on tourism flows and time series models should be analysed in order to come to a proposal for a suitable forecasting tool. In a later stage, more sophisticated models shall be investigated, given that an appropriate data base will be available. Cooperation with academic researchers right from the beginning was considered to be a good way to involve high-level experts and arrange a long-term support also for the future routine-wise operation of the forecasting system.

The results of the forecasting project are reported in two parts:

1. Tourism flow data: Results of the empirical analysis
2. Forecasting methods: Characteristics and comparison

The first part gives a detailed survey of tourism flow data which are in the core of the project: the number of arrivals of visitors and the number of overnights. The presentation contains all relevant details of these time series like the trends and seasonality. Results are also shown for breakdowns for countries of origin of the visitors. Based on the experience of the forecasting project, recommendations to the owner of the data are provided that may help to improve the data base for the future forecasting system.

The second part presents the results from the analysis of forecasting methods. The data base were monthly numbers of arrivals and of overnights. The analysis was focused on two forecasting techniques, Triple Exponential Smoothing and Seasonal Autoregressive Integrated Moving-Average (SARIMA). The evaluation of the methods was done (1) by means of an ex-ante forecasting analysis whereby five different subsets of the data, 2000-2010, 2000-2011, …, and 2000-2014, were used to produce the monthly out-of-sample forecasts of the respective subsequent year, and (2) by calculating out-of-sample forecasts for 2016-2018. The comparison of the methods was based on the forecast errors. The evaluation of the forecast errors revealed that exponential smoothing outperforms SAR(I)MA in most analyzed cases.

The results of the study indicate that a forecasting system based on exponential smoothing is well suited for the implementation of routine-wise forecasting the tourism flows of Greek tourism. Forecasts of the numbers of arrivals and of overnights can be calculated, for inbound and for domestic visitors, and for breakdowns for countries of origin of the visitors.

For the future routine-wise operation of the proposed forecasting methods, an institutional arrangement has to be established. The institution has to deal with the development of an IT system which allows handling the data and estimate the forecasts, has to establish the data base for the forecasting system and care for updating and editing the data, has to run the forecasting procedures, and produce and disseminate the forecasting reports. Agreements with the owners of input data like ELSTAT and cooperation with academic research are recommended.

Future extensions shall focus on forecasts of regional tourism flows. Also, the use of more sophisticated models like Vector Autoregressive models, Error Correction models, Time Varying Parameter (TVP) models, etc. shall be envisaged, aiming at deeper analyses of Greek tourism, in particular the analysis of the economic background of the tourism key statistics.

Four appendices contain results from empirical data analyses and a comprehensive review of literature on tourism forecasting.

# **Introduction**

Forecasts of tourism statistics are of crucial importance and interest for all institutions which are actors in the Greek tourism such as political authorities, tourism businesses, and agencies that are supporting tourism or representing actors in tourism like the Ministry of Tourism, GNTO, the Hellenic Chamber of Hotels, SETE, ITEP, etc., also research institutions including academia. The interest in forecasting results concerns primary variables that describe the tourism flow like the number of visitors (arrivals, overnights), the duration of stay, and the expenditures of the visitors. Deeper analyses refer to details of the tourism flow like the inbound flows from certain countries of residence, flows according to purpose of travel, to regions of destination, and others. Deeper analyses may also refer to the generation of flows as a function of factors like disposable income of visitors, price levels in Greece and alternative destinations, travel costs, exchange rates, Greek marketing expenditure, etc.

**The aims of the project** has to be in line with the availability of data. In a first stage, the focus of the project was on forecasting the tourism flows. For the variables of interest (number of arrivals, number of overnights, expenditures), monthly data as published by ELSTAT should be suitable both in scope and in quality. Modelling approaches involve rather complex methodological issues and shall be postponed at the time being to a later stage.

Among the statistical tools for documenting and forecasting tourism development, the forecasting methodology is the most demanding one. The forecaster is confronted with issues related both to the availability of data which can be used for trend extrapolation and model fitting and to the methods which must be suitable for the forecasting problem. Within the Technical Assistance Action, the Activity I.1.2 has the focus on the development of tools for forecasting the tourism development.

As a first step of the forecasting project, a literature survey was undertaken about methods that have been used for forecasting Greek tourism. The survey was extended to methods as published in the WTO Handbook on Tourism Forecasting Methodologies, in scientific journals, in particular in the Journal of Travel Research, in conference papers and books.

In meetings with representatives from academia, potentials for co-operation in the forecasting task have been investigated. Talks took place with representatives from the University of Piraeus and from the University of Patras. These talks revealed interest of the academics in co-operating in the forecasting task; e.g., research on related methodological issues would be suitable topic for a student or a post-doc research work.

**A cooperation in the forecasting project** has been arranged with Dr. Yiannis Smirlis from the University of Piraeus, and his colleague, Dr. Vangelis Tsioumas from The American College of Greece - DEREE. The cooperation was agreed to include the analysis of the available data and the investigation of suitable forecasting methodologies. The aim of the project was defined to investigate the use of time-series models for the extrapolation of time-series of tourism flow statistics, possibly also expenditures if such data are available. The focus was to investigate the suitability of various forecasting methods for tourism statistics, taking into account the time-series characteristics like trends, seasonality, and structural breaks.

Data basis are tourism statistics as published by ELSTAT but also by other data providers like Eurostat and SETE. ELSTAT publishes monthly tourism flow statistics as well as tourism expenditures that are based on various sources like the Frontier Survey of the Bank of Greece, on the Vacation Survey, the Household Budget Survey, and on the Survey in Hotels, Similar Establishments and Tourist Campsites, all three conducted by ELSTAT, and other data sources.

**Forecasting methods**: It was decided that forecasts of the numbers of arrivals and the numbers of overnights in hotels etc. would be subject of the first phase of the project. Monthly numbers of arrivals and numbers of overnights in hotels etc. are available. The data of at least 20 years should be used for fitting the models. Suitably disaggregated series with respect to national and regional tourism, in-bound and domestic tourism, in-bound tourism flows from certain countries of residence, and others should be used to estimate forecasts. The forecasting horizon should be up to 12 months. As forecasting methods, time-series methods like

* ARIMA and SARIMA
* Exponential smoothing

should be investigated and analysed with respect to forecasting accuracy and other quality criteria.

Methods that are based on more sophisticated models like Vector Autoregressive models, Cointegration and Error Correction models, Time Varying Parameter (TVP) models, etc. should be analysed in the second and advanced stage of the project. Such methods may allow for deeper analyses of Greek tourism, in particular of the economic background of the tourism flows, e.g., the economic determinants of inbound flows from certain countries of residence such as the effect of changes in the disposable income of the visitors, the price levels in Greece and alternative destinations, the travel costs, but also factors of the economic environment like the exchange rates, and factors like the Greek marketing expenditures, events like political unrests, and others.

**Tourism forecasts for politics and business**: For the future routine-wise operation of the then developed forecasting tool, an institutional arrangement has to be established, an IT system has to be developed which allows handling the data, running the forecasting procedures and producing the output that is to be published regularly.

# **The data set. Empirical Analysis and Results**

## **Data collection, data base modelling**

The initial data set was provided by the ELSTAT in Excel files. The data were organized in observations (records), each one of which corresponds to one month, starting from January 2000 and ending to December 2015. The total number of observations is 192.

The variables of the data set (record layout) appear in Figure 2.1-1

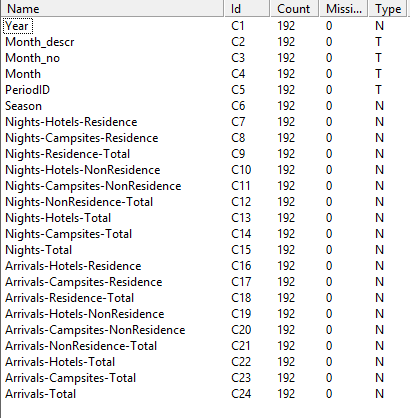


Figure 2.1-1. The variables of the data set

The first five (C1-C5), i.e. Year, Month\_descr, Month\_no, month and PeriodID identify the month and the year of the observation and are used as auxiliary variables. Next variable C6 Season has been estimated to distinguish time observation in three clusters, depending on the amount of tourist arrivals and nights spent. This variable is discussed in detail in Chapter 2. The rest, C7-C24, describe the number of arrivals and nights spent and are include level of analysis for the type of accommodation (Hotels, Campsites) and origin of the tourists (Residents, Non-Residents). They appear in a hierarchical structure, with the extension “-Total” in the name of the variable to denote that the associated variable derives as sum of corresponding variables in the lower of the hierarchy. The hierarchical structure of the data model appears in Figure 2.1-2. Thus, for example the variable Nights-Hotel-Total, derives as the sum of Nights-Hotel-Residents and Nights-Hotel-Non-Residents, i.e. Nights-Hotel-Total = Nights-Hotel-Residents + Nights-Hotel-Non-Residents.

In the upper level, the totals for Nights and Arrivals are estimated by the equations

Nights-Total= Nights-Hotel-Total + Nights-Campsites-Total =

= Nights-Residents-Total + Nights-Non-Residents -Total

Arrivals-Total= Arrivals -Hotel-Total + Arrivals -Campsites-Total

= Arrivals -Residents-Total + Arrivals -Non-Residents -Total

Figure 2.1-2. The hierarchical structure of the data

Next Figure 2.1-3 presents an indicative number of records in the data set.

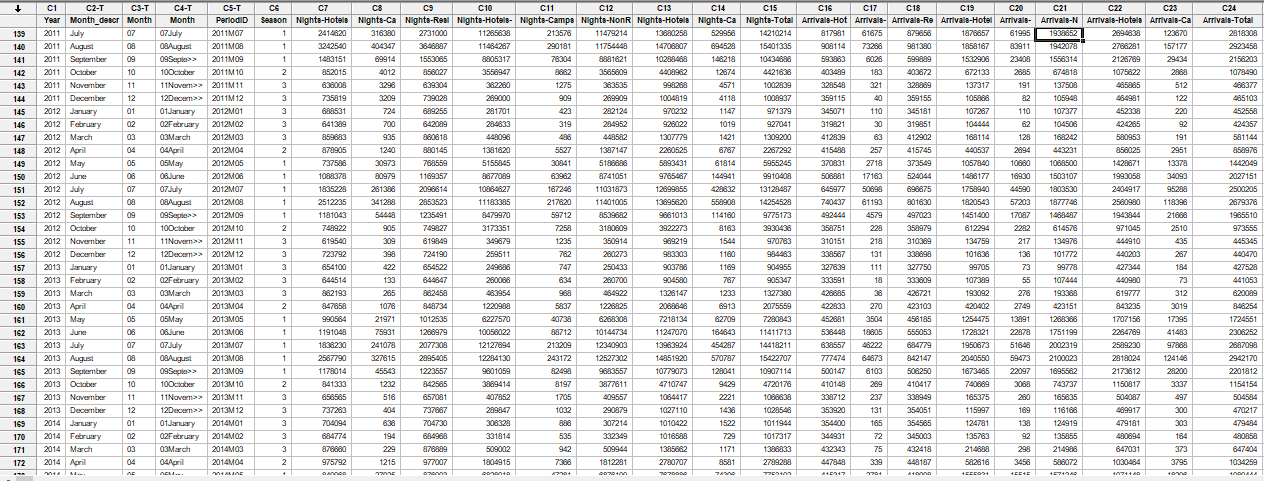


Figure 2.1-3. A sample of records in the data set

## **Descriptive Statistics**

In the Figures included in Appendix I, a short statistical summary is presented for each one of the variables in the data set. This summary includes presentation of the basic statistical measures, the associated histogram, boxplot and 95% interval estimations for the mean and the median.

However, due to the great variability of the data between distinct months of the year, in almost all of the cases, values for high season months (July, August etc.) appear as extreme outliers and affect the reliability of important measurement such as the mean. For this reason, greater attention has been given in times series trend analysis that follow in the next sections.

## **2.3 Analysis by month**

In this section, we investigate the behaviour of the number of arrivals (variable Arrivals-Total) and nights spent (variable Nights-Total) by exploiting time-series graphs and statistics, associated with the months, from January to December. For every month, we provide three charts for the variable Arrivals-Total and three for Nights-Total (see Appendix III). The first chart in each group presents a time-series linear trend estimation for the variable under consideration. The second presents the basic statistics including statistical measures, histogram, boxplot and confidence intervals 95% for the mean and median. The third chart plots the time-series data for the components of the associated variable to explore their contribution.

At a first level of analysis, both numbers of arrivals and nights spent show great variation within each year, ranging between maximum values that correspond to summer months and minimum values in the winter months. Figure 2.3-1 and 2.3-2 show this variation for the Nights-Total and Arrivals-Total, respectively. The same graphs also indicate a positive (increasing) trend, especially from observation 95 (approx. year 2007) and afterwards.



Figure 2.3‑1 Time-series plot for Nights



Figure 2.3‑2 Time-series plot for Arrivals

The boxplots for the months of the year shown in Figures 2.3-3 and 2.3-4, present the same seasonal behaviour of months, as previously mentioned. For both variables, the maximum values and greater variation are observed in months between May-September. The pie charts next to the boxplots, also indicate that the summer months in high touristic season have obviously the greater contribution (percent) to the total number of tourist within the year.

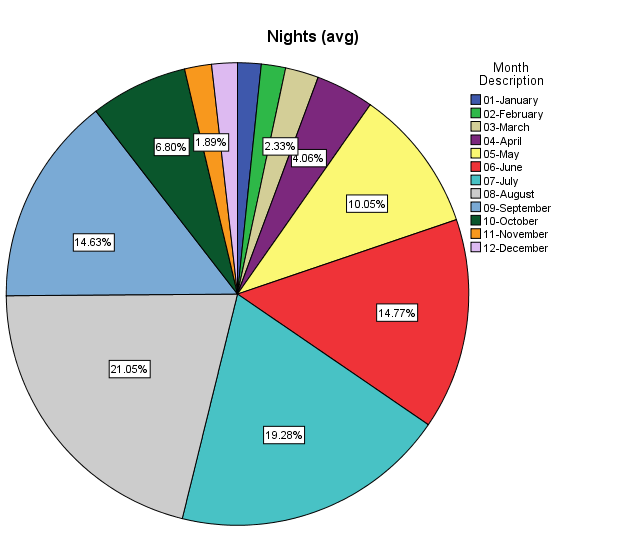


Figure 2.3‑3 Box plot and pie chart for Nights-Total

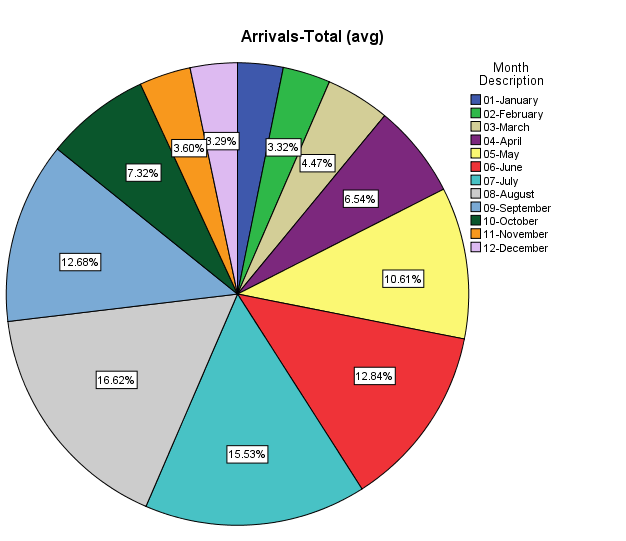


Figure 2.3‑4 Box plot and pie chart for Arrivals-Total

In Appendix II, plots of the main data set variables are presented. From these it derives that:

1. Months December, January, February and March appear with great differences among the years. Particularly, in the period 2004-2010 there is a strong increase is observed of both arrivals and nights. For the next 2-3 years, there is strong decrease which after 2012-13 turns again to increase. In total, in most of the cases, graphs show a slight positive trend without any significant changes. The shape of the variation between the years diverges from linear trend.
2. The number of nights (variable Nights-Total), for months April, March and November, show a negative trend within the years. This behavior is an exception, as for all the rest of months the trend is slightly or strong positive.
3. During the first four months, January to April, and within the period 2003-2004, there is a significant decrease to the number of nights for non-residents. In the opposite direction, the number of nights in Hotels show a significant growth, especially in February and March.
4. For the next months, May to September, the shapes of the time-series graphs change with approximately the same pattern, showing that these months should be considered as a cohesive time-period. The relative graphs for these months indicate a strong positive trend that approaches to linearity, as the reported relative small values for Mean Absolute Deviation, Mean Square Error and Mean Absolute Percentage Error show. The projection of the linear trend line in the future period 2017-2018 gives an indication about what is expected for those months. For example, from Figures II-38 and II-41 derive that the approximated values for July 2017 in Nights-Total will be approx. 16 million and in Arrivals-Total between 3-3,25 millions. In the same manner, for August 2017, Figures II-44 and II-47 give approximations of 17 millions for Nights-Total and of 3.2-3.6 millions for Arrivals-Total.
5. For October and November, the time-series pattern is similar to that of months March and April. Note that in year 2012, the smallest value for both nights and arrivals is observed. After that year follows a period with a strong increase.
6. Concerning the arrivals and nights spent in Campsites, their number is very small and remain constant, relative to that of hotels. Figures for Hotels show a significant positive trend in almost all years of the study period.
7. The figures for residents tourists for both nights and arrivals show a significant step decrease in 2003-2004 after which follows a period of non-significant variations. In an opposite direction, figures for non-residents show an increasing trend.

## **2.4. Analysis by Seasons**

A basic conclusion extracted from the time-series analysis presented in the previous section was that months can be grouped in three similar clusters in terms of touristic activity, particularly by considering the variables Nights-Total and Arrivals-Total. The analysis of the previous chapter indicated that these three classes may be as follows :

Class I – Low Season : November, December, January, February, March

Class II – Intermediate Season : October, April

Class III – High Season : May, June, July, August, September

In this chapter we provide further documentation to support such a month clustering and further investigate the distinct trends of these clusters.

### **2.4.1. Definition of Seasons**

In order to explore the above defined grouping of months, a new variable Season has been introduced in the data set. This variable accepts values 1, 2 and 3, depending on which season the particular time observation belongs.

Based on these values, the mean and Anova tests presented at Appendix III, show statistically significant difference of the mean values in the three seasons of the variables Nights-Total and Arrivals-Total and thus confirms the distinction between them. The boxplot graphs in the following Figures 2.4-1, 2.4-2 provide a further indication about the differences between the three seasons.



Figures 2.4‑1 Boxplot of Nights-Total by Month



Figures 2.4‑2 Boxplot of Arrivals-Total by Month

### **2.4.2. Analysis of Seasons**

Table 2.4-1 presents simple descriptive statistics for Arrivals-Total and Nights-Total when the classification based on the variable Season is considered. The range and the distribution of values within the classes is also presented graphically in the boxplots and histograms in Figures 2.4-3 - 2.4-6.

The obtained classification of the months applies to the rest of the variables of the data set. In Figures 2.4-7, 2.4-8, 2.4-9 and 2.4-10 interval graphs present a clear distinction of values when the seasons 1, 2 and 3 are related to the variables Nights-Residents-Total, Nights-Non-Residents-Total, Nights-Campsites-Total, Nights-Hotel-Total and Arrivals-Residents-Total, Arrivals-Non-Residents-Total, Arrivals-Campsites-Total, Arrivals-Hotel-Total.

Concerning the time-series trend when classes are considered, Figures 2.4-11 and 2.4-12 present composite trend lines for Nights-Total and Arrivals-Total. In Figures 2.4-11 for Nights-Total , classes 2 and 3 show a constant trend within the years but class 1 (high season) appears with a slight increasing trend especially after observation number 57 which corresponds to year 2004. This trend is evident in Figures 2.4-12 which plots Arrivals-Total.

Next, follow the Tables and Figures mentioned in this section.

Table 2.4-1. Simple descriptive statistics of Arrivals-Total and Nights-Total by Season

Variable Season Mean StDev Minimum Maximum

Arrivals-Total 1 2050593 485580 1285523 3247472

2 1040318 118331 846254 1351422

3 536555 87233 395040 770093

Nights-Total 1 10158424 2912383 5418574 16527563

2 3538000 1024726 2075559 5288263

3 1142992 179382 828899 1575610



Figures 2.4‑3 Boxplot of Nights-Total by Season



Figures 2.4‑4 Boxplot of Arrivals-Total by season.



Figures 2.4‑5 Histograms for Arrivals-Total by Season.



Figures 2.4‑6 Histograms for Nights-Total by Season.



Figures 2.4‑7 Interval plot for Nights-Residents-Total and Nights-Non-Residents-Total by Season.



Figures 2.4‑8 Interval plot for Arrivals-Residents-Total and Arrivals-Non-Residents-Total by Season.



Figures 2.4‑9 Interval plot for Arrivals-Campsites-Total and Arrivals-Hotels-Total by Season.



Figures 2.4‑10 Interval plot for Nights-Campsites-Total and Nights-Hotels-Total by Season.



Figures 2.4‑11 Time-series composite graph for Nights-Total by Seasons



Figures 2.4‑12 Time-series composite graph for Arrivals-Total by Seasons

## **2.5. Correlations**

This section investigates possible correlations between nights spend and number of arrivals. From the analysis, the following results and comments derive.

1. The variable Nights-Total, considered as dependent variable, is linearly correlated to Arrivals-Total with a high significant degree (R2=98%). Figure 2.5-1 and the following regression analysis document this observation.
2. However, this strong relation between total values for arrivals and nights spent is not applied to other breakdown factors such as Campsites, Resident and NonResident. Scatterplot graphs presented in Figures 2.5-2, 2.5-3 and 2.5-4 indicate this independent relation.
3. When examine the relation between Nights-Hotels-Total and Arrivals-Hotels-Total, the same strong correlation with the corresponding total values is observed (Figure 2.5-5). This is due to the fact that Campsites-Total, being the second factor, has very small values relative to those of Hotels-Total.



Figure 2.5‑1 Scatterplot for Nights-Total vs Arrivals-Total

Regression Analysis: Nights-Total versus Arrivals-Total

The regression equation is Nights-Total = - 2446610 + 5,90 Arrivals-Total

Predictor Coef SE Coef T P

Constant -2446610 88111 -27,77 0,000

Arrivals-Total 5,89679 0,06002 98,24 0,000

S = 638250 R-Sq = 98,1% R-Sq(adj) = 98,1%

PRESS = 7,897867E+13 R-Sq(pred) = 98,03%



Figure 2.5‑2 Scatterplot for Nights-Campsites-Total vs Arrivals-Campsites-Total



Figure 2.5-3 Scatterplot for Nights-Non-Residents-Total vs Arrivals- Non-Residents-Total



Figure 2.5-4 Scatterplot for Nights-Residents-Total vs Arrivals-Residents-Total



Figure 2.5-5 Scatterplot for Nights-Hotels-Total vs Arrivals-Hotel-Total

The regression equation is

Nights-Hotels-Total = - 2482278 + 5,978 Arrivals-Hotels-Total

R-Sq(pred) = 97,9%

## **2.6. Analysis of Arrivals by country of origin**

This section presents descriptive statistics and time-series trend analysis on data measuring the tourist arrivals by country of origin. The data have been retrieved from the ELSTAT web site ([www.elstat.gr](http://www.elstat.gr)) and refer to years 2007-2015. They were published in the form of MS-Excel files named “A2001\_STO04\_TB\_QQ\_00\_<year>\_01\_F\_GR and were structured in summary tables. Data for other factors such as nights spent, types of accommodation etc. were not available on the ELSTAT website.

The structure of this report is as follows: Section 2.6.1 presents the collected data set and the basic descriptive statistics measurements. Section 2.6.2 focuses on the most important countries in terms of their number of arrivals and examines the total number in years 2007-2015. Section 2.6.3 presents plots the monthly time-series for each country and estimates its trend.

### **2.6.1. The data set**

The data is organized in observations (records), each one of which corresponds to one month, starting from January 2007 and ending on December 2015. The total number of countries mentioned is 48 and the total number of observations is 5015.

Note that in the published tables there was a reference for “Rest Asian, African… countries”. The associated data were not included in this analysis. Moreover, Serbia was reported sometimes as Serbia and others as Serbia-Montenegro and for simplicity reasons the two cases were joined in one.

The data set has the following structure:

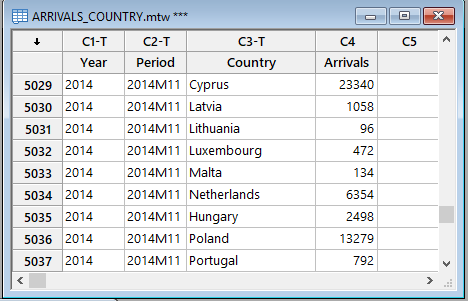
YEAR = year of the observation

PERIOD = in the form <YEAR>M<Month>. E.g., 2014M11 is for November 2011

COUNTRY = Name of the country of origin

ARRIVALS = Number of arrivals recorded

A sample of few records is presented in the following figure.



The associated table for the basic descriptive statistics are presented in the following Table 2.6-1.

Table 2.6-1. Descriptive statistics of arrivals by country of origin

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | Mean | St.Dev | Q1 | Median | Q3 | Min | Max | Range | N |
| Albania | 29663 | 13561 | 17439,5 | 30009 | 39409 | 541 | 64513 | 63972 | 108 |
| Argentina | 1981 | 2453 | 457,5 | 1562 | 2687 | 0 | 16459 | 16459 | 108 |
| Austria | 26094 | 27435 | 3286,8 | 12210 | 51471 | 1172 | 91081 | 89909 | 108 |
| Belgium | 32401 | 32954 | 4729,5 | 19104 | 58279 | 641 | 119338 | 118697 | 108 |
| Brazil | 3123 | 2896 | 888,3 | 2344 | 4494 | 0 | 13519 | 13519 | 108 |
| Bulgaria | 74621 | 51057 | 44191,0 | 56748 | 75982 | 16885 | 290083 | 273198 | 108 |
| Canada | 12484 | 10507 | 3135,5 | 11091 | 19023 | 0 | 51407 | 51407 | 108 |
| China | 1807 | 1826 | 404,0 | 1203 | 2791 | 0 | 7725 | 7725 | 108 |
| Croatian | 1345 | 1634 | 0,0 | 607 | 2477 | 0 | 5350 | 5350 | 34 |
| Cyprus | 38005 | 15851 | 27038,8 | 35950 | 46821 | 0 | 83130 | 83130 | 108 |
| Czech republic | 25662 | 32834 | 1298,5 | 4213 | 52882 | 0 | 127742 | 127742 | 108 |
| Denmark | 19898 | 21441 | 1433,5 | 13162 | 35887 | 0 | 79932 | 79932 | 108 |
| Egypt - Sudan | 1018 | 1189 | 0,0 | 732 | 1411 | 0 | 6382 | 6382 | 108 |
| Estonia | 1541 | 1969 | 0,0 | 688 | 2518 | 0 | 8383 | 8383 | 108 |
| Finland | 13653 | 13279 | 895,8 | 11387 | 25078 | 0 | 54459 | 54459 | 108 |
| France | 92557 | 90459 | 13396,8 | 66750 | 154266 | 5439 | 388076 | 382637 | 108 |
| FYROM | 251922 | 254650 | 69897,0 | 108666 | 480637 | 63853 | 725789 | 661936 | 12 |
| Germany | 198801 | 166347 | 41863,5 | 127985 | 336953 | 19094 | 558146 | 539052 | 108 |
| Hungary | 9524 | 12122 | 1578,3 | 4057 | 13797 | 195 | 59776 | 59581 | 108 |
| Iceland | 245 | 601 | 0,0 | 0 | 297 | 0 | 4700 | 4700 | 108 |
| Iran | 478 | 945 | 0,0 | 0 | 536 | 0 | 5857 | 5857 | 108 |
| Ireland | 5424 | 6138 | 816,8 | 2099 | 9659 | 0 | 24510 | 24510 | 108 |
| Israel | 12880 | 13811 | 2629,0 | 6912 | 18673 | 0 | 67847 | 67847 | 108 |
| Italy | 86612 | 107673 | 19893,0 | 34597 | 119968 | 9890 | 500396 | 490506 | 108 |
| Japan | 2388 | 9495 | 353,3 | 897 | 1535 | 0 | 73666 | 73666 | 108 |
| Latvia | 2220 | 4137 | 0,0 | 882 | 3049 | 0 | 29314 | 29314 | 108 |
| Lebanon - Syria | 1473 | 1738 | 284,8 | 1047 | 2090 | 0 | 10384 | 10384 | 108 |
| Lithuania | 2687 | 3475 | 187,0 | 1215 | 4045 | 0 | 19909 | 19909 | 108 |
| Luxembourg | 1770 | 2362 | 191,0 | 879 | 2217 | 0 | 13608 | 13608 | 108 |
| Malta | 613 | 1592 | 0,0 | 0 | 601 | 0 | 11735 | 11735 | 108 |
| Mexico | 1011 | 1033 | 0,0 | 754 | 1691 | 0 | 5072 | 5072 | 108 |
| Netherlands | 51646 | 50354 | 6351,0 | 35562 | 93574 | 0 | 191506 | 191506 | 108 |
| Norway | 21113 | 22633 | 1117,3 | 8762 | 35597 | 0 | 75382 | 75382 | 108 |
| Poland | 32715 | 43465 | 3346,3 | 9326 | 50147 | 458 | 184920 | 184462 | 108 |
| Portugal | 1682 | 2001 | 405,0 | 958 | 2193 | 0 | 12488 | 12488 | 108 |
| Romania | 27902 | 22579 | 12213,3 | 20805 | 35445 | 149 | 118643 | 118494 | 108 |
| Russia | 55134 | 76962 | 4033,8 | 13997 | 78700 | 0 | 320577 | 320577 | 108 |
| Serbia - Montenegro | 57845 | 75751 | 9175,5 | 25557 | 85002 | 1702 | 339655 | 337953 | 108 |
| Slovakia | 5065 | 7687 | 280,5 | 990 | 8435 | 0 | 32608 | 32608 | 108 |
| Slovenia | 3063 | 5182 | 272,3 | 867 | 3696 | 0 | 33247 | 33247 | 108 |
| South African Union | 2024 | 2322 | 472,8 | 1143 | 2853 | 0 | 12711 | 12711 | 108 |
| South Korea | 731 | 871 | 15,0 | 444 | 1207 | 0 | 5870 | 5870 | 108 |
| Spain | 12543 | 12030 | 4505,8 | 7813 | 16600 | 1246 | 55705 | 54459 | 108 |
| Sweden | 28152 | 27173 | 2656,8 | 21489 | 51104 | 0 | 93756 | 93756 | 108 |
| Switzerland | 29045 | 25686 | 5332,0 | 20573 | 50409 | 1782 | 92493 | 90711 | 108 |
| Turkey | 47690 | 44380 | 16421,0 | 39432 | 66093 | 0 | 251973 | 251973 | 108 |
| United Kingdom | 173268 | 160740 | 26697,0 | 120160 | 327008 | 11475 | 538093 | 526618 | 108 |
| USA. | 45396 | 31111 | 18150,3 | 40085 | 66942 | 753 | 135079 | 134326 | 108 |

From Table 2.6-1 derives that Croatia and FYROM have limited number of observations (34 and 12 respectively), as the last column (labeled N) indicates. Moreover, the minimum number of arrivals in many cases is zero, denoting no touristic activity. This happens in winter and low season months.

### **2.6.2. Most significant countries**

To identify the most significant countries, the total number of arrivals per year has been calculated and based on that the share (percentage) of each, relative to the total arrivals for all countries has been extracted.

Table 2.6-2 presents the figures of the total number of arrivals in years 2005-2017 for the 48 countries included in the analysis. The two last columns present the total number of arrivals in years 2005-2017 and the percentage relative to the grand total.

Note that FYROM appears to have reported values for only 2015 and the total number is high enough to result in a percentage of 2,11%. However, due to the limited data, FYROM has not been included in the analysis.

In order to further examine the behavior of each country, the most significant of them have been identified. To distinguish them, an empirical rule has been applied: The countries that exceeded the threshold of 2% were identified as the most significant countries. The countries that have such level of arrivals are: Germany, United Kingdom, France, Italy, Bulgaria, Serbia – Montenegro, Russia, Netherlands, Turkey, USA, Cyprus, Poland, Belgium, Albania, Switzerland, Sweden, FYROM, and Romania. Their total share reaches 85,3% of the total arrivals in Greece for the period 2005-2017. These countries are listed in Table 2.6-3. Figure 2.6-1 presents graphically the total number of arrivals and the percentage of each country.

Germany appears to be the most significant country with a percentage of 15,01%. It is followed by United Kingdom with 13,09%, France with 6,99% and Italy with 6,54%.

Table 2.6-2. Total arrivals by Country of Origin in years 2007-2015

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | Total | Percent |
| Albania | 213725 | 242999 | 234276 | 242083 | 411245 | 469215 | 456391 | 442325 | 491380 | 3203639 | 2,24% |
| Argentina | 9390 | 14299 | 13878 | 17772 | 13834 | 20207 | 41947 | 47923 | 34662 | 213912 | 0,15% |
| Austria | 377341 | 354747 | 352223 | 338367 | 310357 | 236417 | 236475 | 285131 | 327121 | 2818179 | 1,97% |
| Belgium | 408655 | 420747 | 334240 | 339836 | 432625 | 326938 | 344554 | 409199 | 482524 | 3499318 | 2,45% |
| Brazil | 23534 | 27487 | 24540 | 34015 | 52119 | 31125 | 26673 | 52220 | 65589 | 337302 | 0,24% |
| Bulgaria | 701666 | 623476 | 657130 | 664389 | 686208 | 599109 | 691873 | 1534565 | 1900643 | 8059059 | 5,64% |
| Canada | 158795 | 158447 | 134983 | 113358 | 142287 | 102694 | 198543 | 156892 | 182299 | 1348298 | 0,94% |
| China | 5331 | 5941 | 7793 | 13620 | 15838 | 12203 | 29335 | 49943 | 55098 | 195102 | 0,14% |
| Croatian |  |  |  |  |  |  | 3686 | 21891 | 20137 | 45714 | 0,03% |
| Cyprus | 492473 | 474942 | 434746 | 574764 | 439757 | 424827 | 373875 | 419029 | 470092 | 4104505 | 2,87% |
| Czech republic | 269774 | 267597 | 267833 | 294936 | 309061 | 289036 | 287801 | 348704 | 436704 | 2771446 | 1,94% |
| Denmark | 267650 | 245948 | 264040 | 240563 | 244986 | 205194 | 202476 | 240418 | 237656 | 2148931 | 1,50% |
| Egypt - Sudan | 14645 | 13558 | 12610 | 15925 | 4675 | 4724 | 5558 | 12137 | 26077 | 109909 | 0,08% |
| Estonia | 28732 | 26018 | 21242 | 13842 | 9862 | 4757 | 8093 | 30552 | 23323 | 166421 | 0,12% |
| Finland | 174202 | 147746 | 170341 | 205282 | 167633 | 154134 | 139685 | 165601 | 149893 | 1474517 | 1,03% |
| France | 991118 | 910021 | 962435 | 868346 | 1149389 | 977375 | 1152217 | 1463159 | 1522100 | 9996160 | 6,99% |
| FYROM |  |  |  |  |  |  |  |  | 3023058 | 3023058 | 2,11% |
| Germany | 2711663 | 2469151 | 2364486 | 2038871 | 2240480 | 2108787 | 2267545 | 2459228 | 2810349 | 21470560 | 15,01% |
| Hungary | 201703 | 180914 | 70894 | 109160 | 69756 | 69790 | 83496 | 96505 | 146379 | 1028597 | 0,72% |
| Iceland | 8236 | 4475 | 3340 | 0 | 0 | 2059 | 1995 | 4697 | 1705 | 26507 | 0,02% |
| Iran | 641 | 847 | 1647 | 9189 | 8748 | 13631 | 5491 | 879 | 10599 | 51672 | 0,04% |
| Ireland | 77451 | 93536 | 73167 | 65623 | 58939 | 32358 | 42577 | 69532 | 72653 | 585836 | 0,41% |
| Israel | 73283 | 84221 | 82443 | 197159 | 226111 | 207711 | 209824 | 194395 | 115868 | 1391015 | 0,97% |
| Italy | 1251779 | 1099983 | 935011 | 843613 | 938231 | 848073 | 964314 | 1117711 | 1355329 | 9354044 | 6,54% |
| Japan | 28780 | 10927 | 6765 | 10021 | 10124 | 8841 | 86559 | 85860 | 9982 | 257859 | 0,18% |
| Latvia | 11862 | 31010 | 12027 | 21948 | 7846 | 15301 | 45665 | 60399 | 33709 | 239767 | 0,17% |
| Lebanon - Syria | 14010 | 11458 | 14753 | 4639 | 4916 | 12844 | 36195 | 30002 | 30257 | 159074 | 0,11% |
| Lithuania | 16399 | 34537 | 37501 | 16295 | 13666 | 21600 | 32195 | 57892 | 60143 | 290228 | 0,20% |
| Luxembourg | 20146 | 25136 | 22792 | 18593 | 28475 | 15193 | 20372 | 18449 | 22038 | 191194 | 0,13% |
| Malta | 21745 | 11701 | 4367 | 9651 | 1368 | 2206 | 1619 | 3702 | 9867 | 66226 | 0,05% |
| Mexico | 12010 | 8878 | 8909 | 10470 | 5531 | 8067 | 19791 | 19627 | 15900 | 109183 | 0,08% |
| Netherlands | 737771 | 756939 | 651440 | 528157 | 560723 | 478482 | 577658 | 647509 | 639107 | 5577786 | 3,90% |
| Norway | 213350 | 277304 | 315595 | 187319 | 226627 | 294113 | 271019 | 250039 | 244859 | 2280225 | 1,59% |
| Poland | 227363 | 270039 | 203487 | 402170 | 450617 | 254683 | 382821 | 587620 | 754401 | 3533201 | 2,47% |
| Portugal | 17947 | 24678 | 13300 | 19497 | 34642 | 20484 | 16267 | 16972 | 17821 | 181608 | 0,13% |
| Romania | 350723 | 327261 | 307596 | 257939 | 223699 | 230396 | 263639 | 511860 | 540289 | 3013402 | 2,11% |
| Russia | 199591 | 309072 | 276021 | 451239 | 738926 | 874787 | 1348077 | 1243926 | 512789 | 5954428 | 4,16% |
| Serbia - Montenegro | 553500 | 686996 | 498356 | 706635 | 692059 | 620450 | 774887 | 986524 | 727831 | 6247238 | 4,37% |
| Slovakia | 68643 | 63095 | 48549 | 49406 | 41788 | 44783 | 61478 | 94043 | 75187 | 546972 | 0,38% |
| Slovenia | 33038 | 75186 | 45924 | 40082 | 31130 | 35720 | 18691 | 26700 | 24379 | 330850 | 0,23% |
| South African Union | 34243 | 35688 | 20539 | 19985 | 21982 | 19686 | 17084 | 23278 | 26055 | 218540 | 0,15% |
| South Korea | 9677 | 4391 | 5123 | 7623 | 1846 | 6100 | 8585 | 15526 | 20081 | 78952 | 0,06% |
| Spain | 182644 | 219917 | 164461 | 155302 | 154773 | 155722 | 91988 | 136230 | 93623 | 1354660 | 0,95% |
| Sweden | 311358 | 382922 | 356154 | 281069 | 333906 | 319756 | 367288 | 336383 | 351573 | 3040409 | 2,13% |
| Switzerland | 310295 | 339809 | 352514 | 274418 | 361405 | 299620 | 388853 | 418733 | 391247 | 3136894 | 2,19% |
| Turkey | 182490 | 207609 | 200348 | 561198 | 552091 | 602305 | 770512 | 920962 | 1153046 | 5150561 | 3,60% |
| United Kingdom | 2508651 | 2278015 | 2112149 | 1802203 | 1758092 | 1920794 | 1846332 | 2089529 | 2397169 | 18712934 | 13,09% |
| USA. | 617478 | 612826 | 531276 | 498301 | 484707 | 373832 | 454071 | 580031 | 750251 | 4902773 | 3,43% |

Table 2.6-3. Countries with total arrivals over 2% of the total in years 2007-2015

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | Total | Percent | CV |
| Germany | 2711663 | 2469151 | 2364486 | 2038871 | 2240480 | 2108787 | 2267545 | 2459228 | 2810349 | 21470560 | 15,01% | 0,11 |
| United Kingdom | 2508651 | 2278015 | 2112149 | 1802203 | 1758092 | 1920794 | 1846332 | 2089529 | 2397169 | 18712934 | 13,09% | 0,13 |
| France | 991118 | 910021 | 962435 | 868346 | 1149389 | 977375 | 1152217 | 1463159 | 1522100 | 9996160 | 6,99% | 0,21 |
| Italy | 1251779 | 1099983 | 935011 | 843613 | 938231 | 848073 | 964314 | 1117711 | 1355329 | 9354044 | 6,54% | 0,17 |
| Bulgaria | 701666 | 623476 | 657130 | 664389 | 686208 | 599109 | 691873 | 1534565 | 1900643 | 8059059 | 5,64% | 0,53 |
| Serbia - Montenegro | 553500 | 686996 | 498356 | 706635 | 692059 | 620450 | 774887 | 986524 | 727831 | 6247238 | 4,37% | 0,20 |
| Russia | 199591 | 309072 | 276021 | 451239 | 738926 | 874787 | 1348077 | 1243926 | 512789 | 5954428 | 4,16% | 0,64 |
| Netherlands | 737771 | 756939 | 651440 | 528157 | 560723 | 478482 | 577658 | 647509 | 639107 | 5577786 | 3,90% | 0,15 |
| Turkey | 182490 | 207609 | 200348 | 561198 | 552091 | 602305 | 770512 | 920962 | 1153046 | 5150561 | 3,60% | 0,59 |
| USA. | 617478 | 612826 | 531276 | 498301 | 484707 | 373832 | 454071 | 580031 | 750251 | 4902773 | 3,43% | 0,20 |
| Cyprus | 492473 | 474942 | 434746 | 574764 | 439757 | 424827 | 373875 | 419029 | 470092 | 4104505 | 2,87% | 0,12 |
| Poland | 227363 | 270039 | 203487 | 402170 | 450617 | 254683 | 382821 | 587620 | 754401 | 3533201 | 2,47% | 0,47 |
| Belgium | 408655 | 420747 | 334240 | 339836 | 432625 | 326938 | 344554 | 409199 | 482524 | 3499318 | 2,45% | 0,14 |
| Albania | 213725 | 242999 | 234276 | 242083 | 411245 | 469215 | 456391 | 442325 | 491380 | 3203639 | 2,24% | 0,33 |
| Switzerland | 310295 | 339809 | 352514 | 274418 | 361405 | 299620 | 388853 | 418733 | 391247 | 3136894 | 2,19% | 0,14 |
| Sweden | 311358 | 382922 | 356154 | 281069 | 333906 | 319756 | 367288 | 336383 | 351573 | 3040409 | 2,13% | 0,09 |
| FYROM |  |  |  |  |  |  |  |  | 3023058 | 3023058 | 2,11% |  |
| Romania | 350723 | 327261 | 307596 | 257939 | 223699 | 230396 | 263639 | 511860 | 540289 | 3013402 | 2,11% | 0,35 |
|  |  |  |  |  |  |  |  |  |  | **TOTAL** | **85,30%** |  |

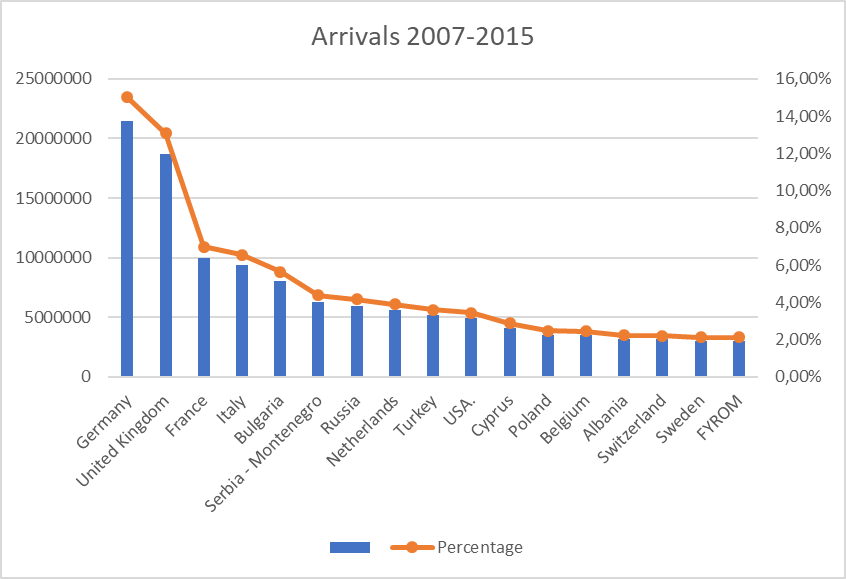


Figure 2.6-1. Total arrivals and percentage of the most significant countries.

In order to further examine how significant are the changes of arrivals within the years of reference in the countries of Table 2.6-3, the coefficient of variation (CV) for each country is estimated and reported in the last column of Table 2.6-3. From this derives that a number of countries have low variation, relatively constant arrival activity and others appear with significantly higher-lower values during the period of the analysis. In the first case (Group 1) there are those countries with CV ≤ 0,25 and in the second case those that have CV ≥ 0,25 (Group 2). Figures 2 and 3 present graphically the variation of total arrivals in years 2005-2017 for group 1 and 2 as previously defined.

It is interesting to note that the majority of EU countries, such as Germany, UK, France, Italy, Sweden, Switzerland, Netherlands, Belgium, and additionally Cyprus, Serbia-Montenegro and USA belong to the first group. These countries have a relatively constant flow of arrivals and this fact indicates that a forecast in year time interval could be possible.

The second group of countries with high variation of arrivals involves mainly Balkan countries (Bulgaria, Albania, Turkey) and Poland and Russia. Particularly for Bulgaria, Figure 2.6-3 shows that it doubled its touristic activity in years 2014 and 2015. On the other side, Russia had a peak in 2013 and after that a great decrease occurred.

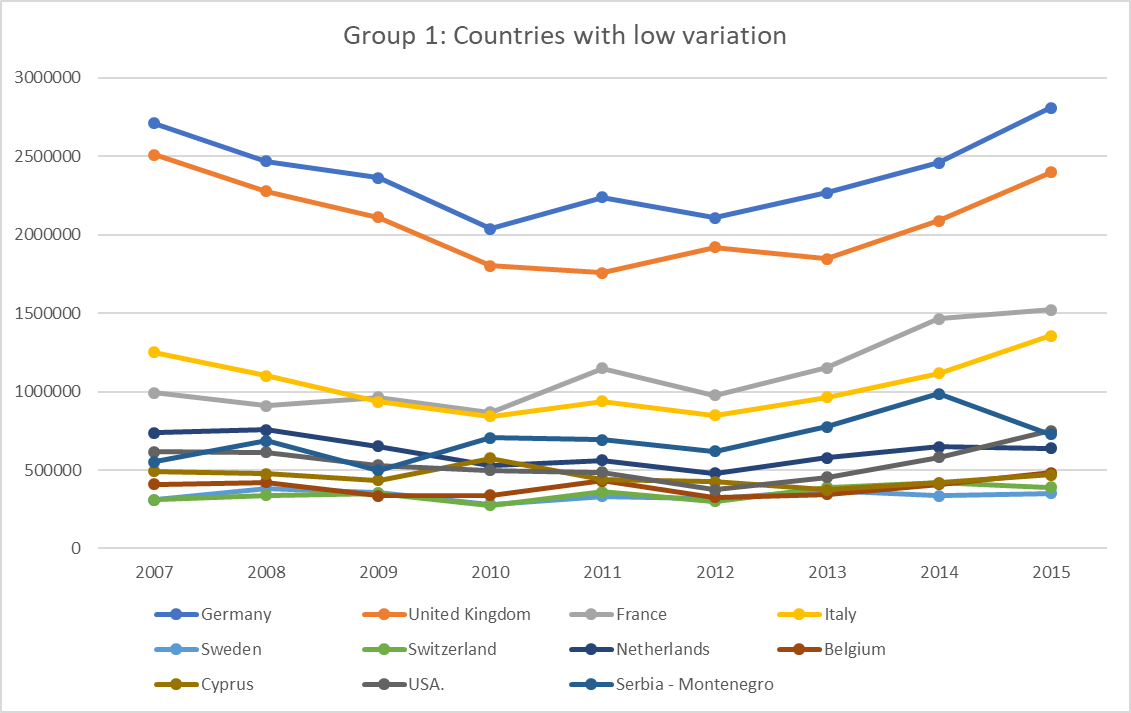
**

Figure 2.6-2. Countries with low arrival variation

Figure 2.6-3. Countries with high arrival variation

### **2.6.3. Presentation of Time-series and Trend**

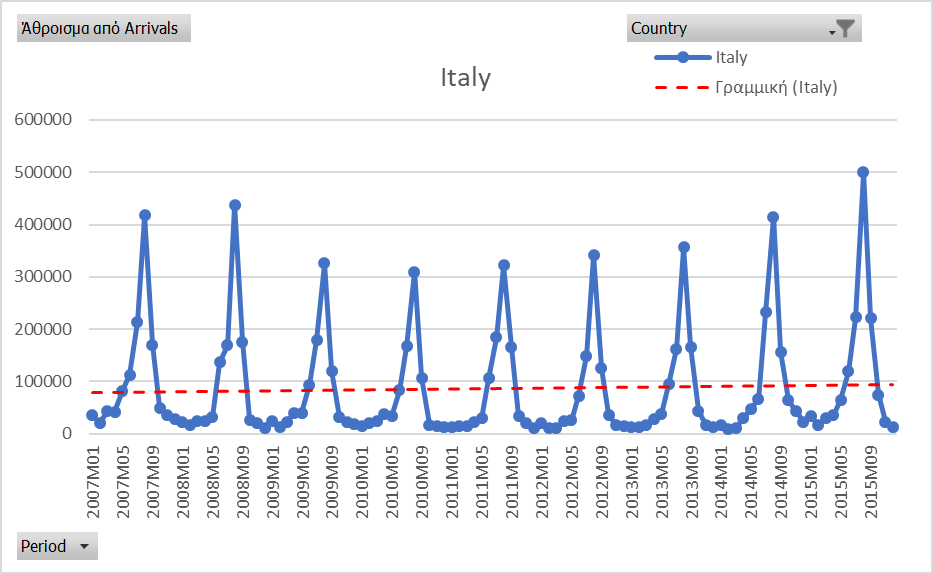
In this section, a time-series presentation of the arrivals for the most important countries is given. In each graph, the trend line is estimated and based on that, the countries have been divided in those that have an almost zero trend, in those that have a significantly positive trend, and those that appear with negative trend.

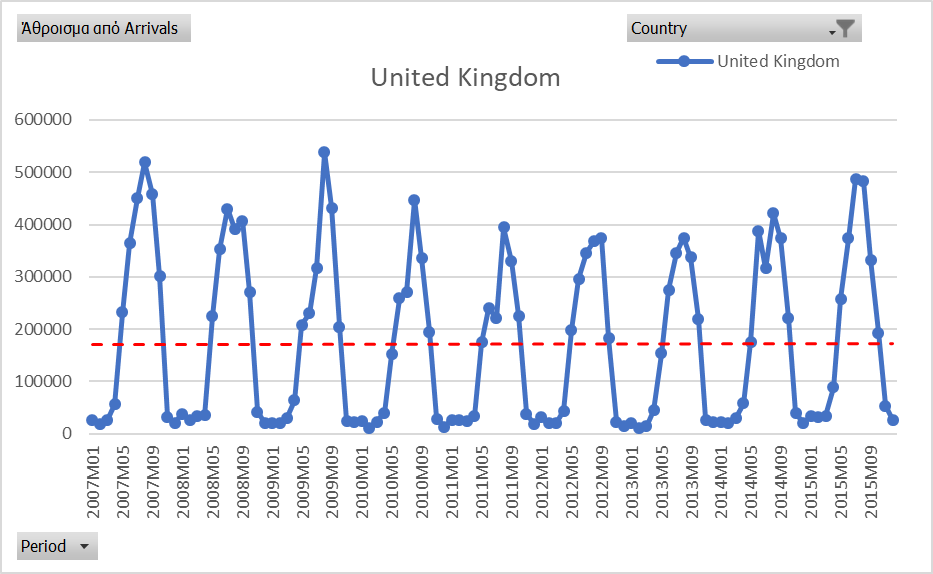
The first group comprises Italy, UK, Germany, Netherlands, USA, Belgium, and Sweden. The group of countries with positive trend involves Russia, Poland, Albania, France, Turkey, Switzerland, Serbia-Montenegro, Romania and Bulgaria. Only Cyprus appears with a relative negative trend.

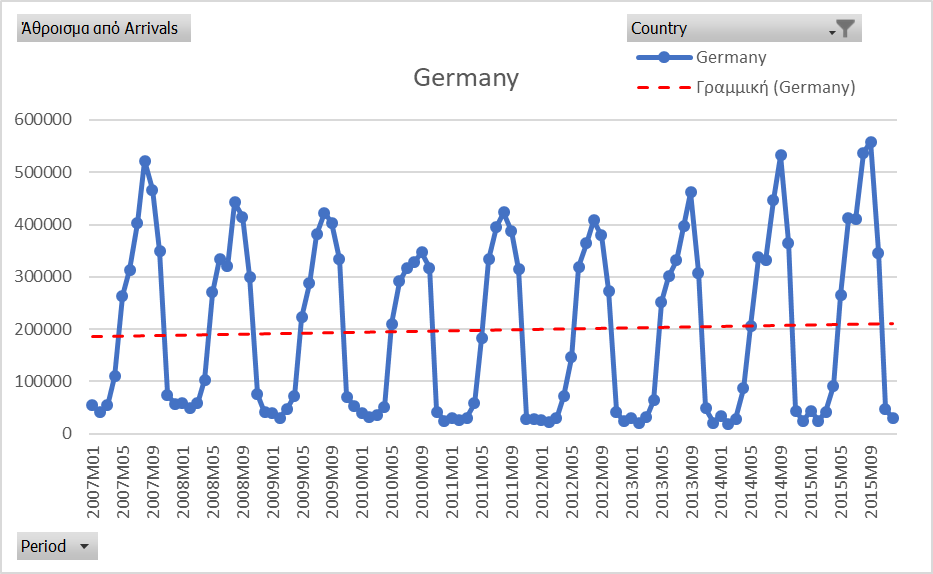
It is important to note that the time-series pattern varies significantly between countries. For example, Albania, Bulgaria, and Turkey do not show the clear distinction between seasons that has been observed for the total number of arrivals. Furthermore, it is interesting to observe that the maximum peak in high season months does not always occur in August as expected. For example, for Germany the maximum number of arrivals is observed in September and for USA this point varies between July, August, and September, depending on the year of reference.

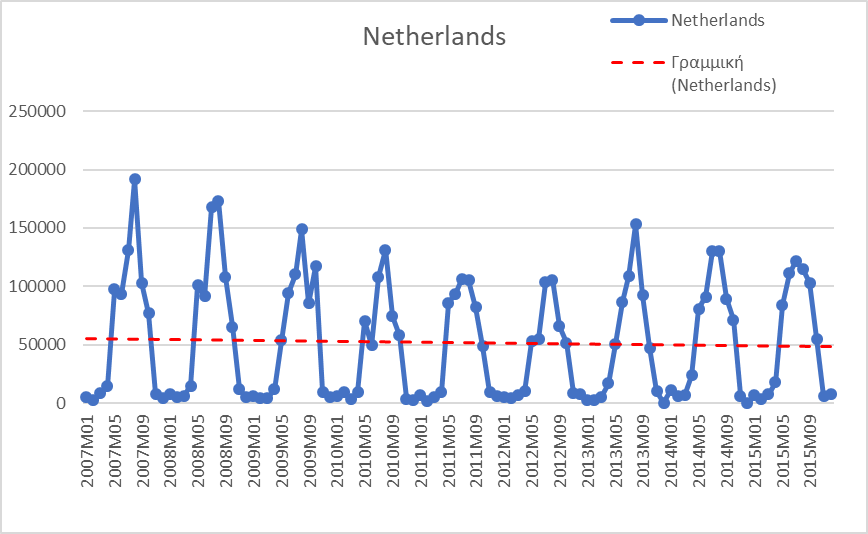
Below are the relevant graphs for each country.

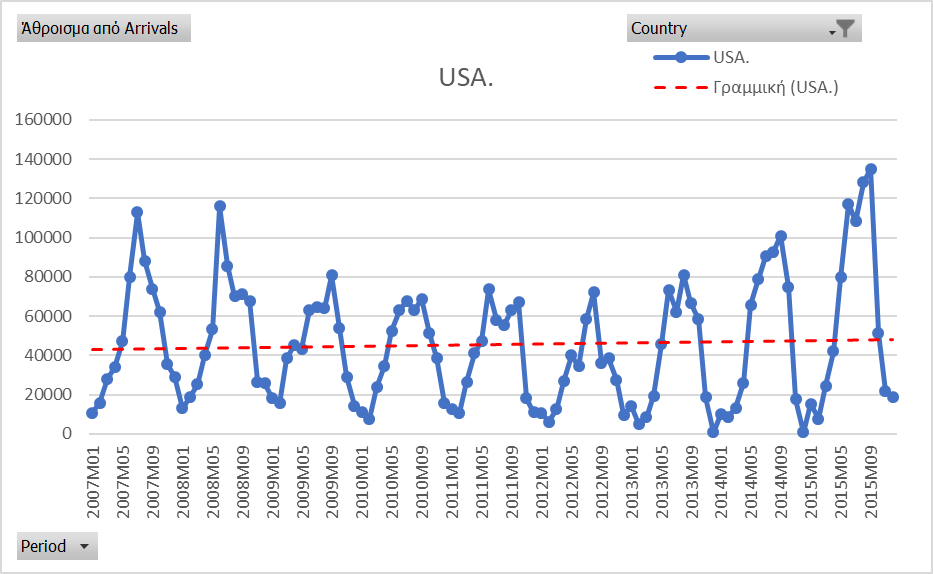
### Countries with almost zero trend

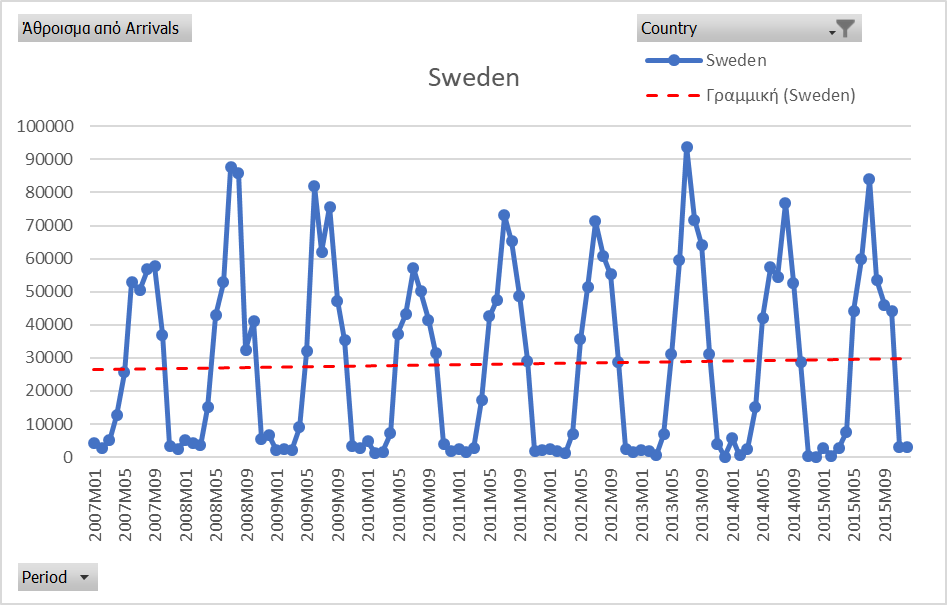


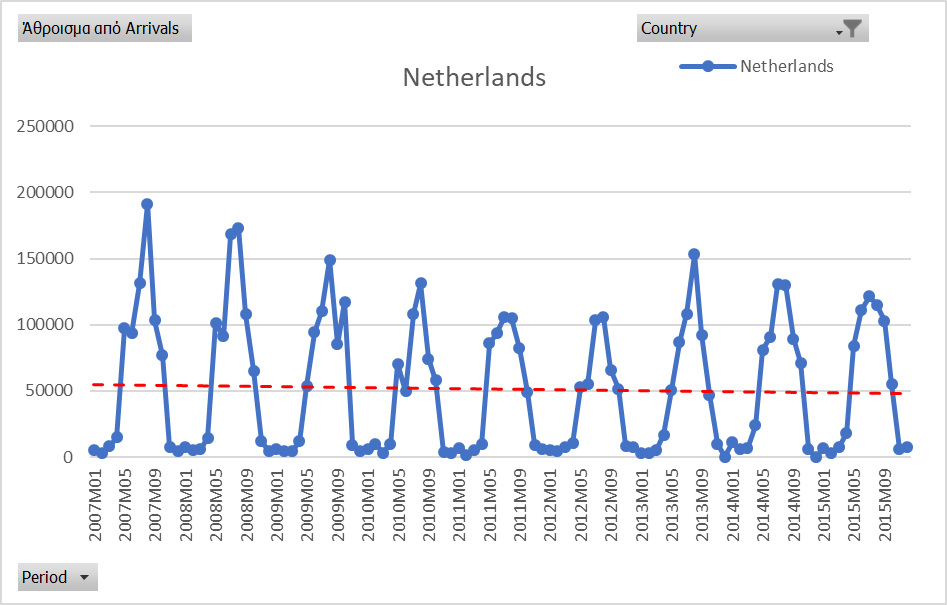




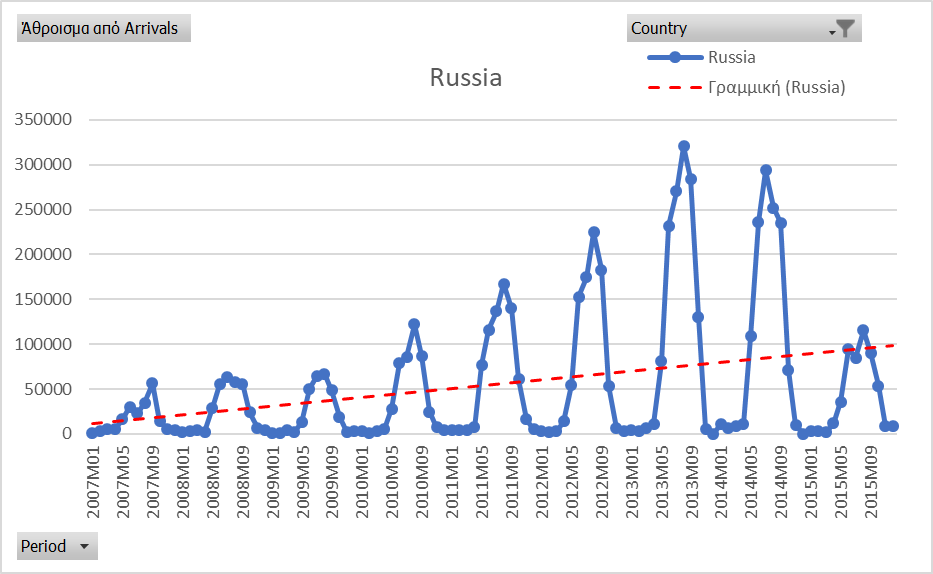


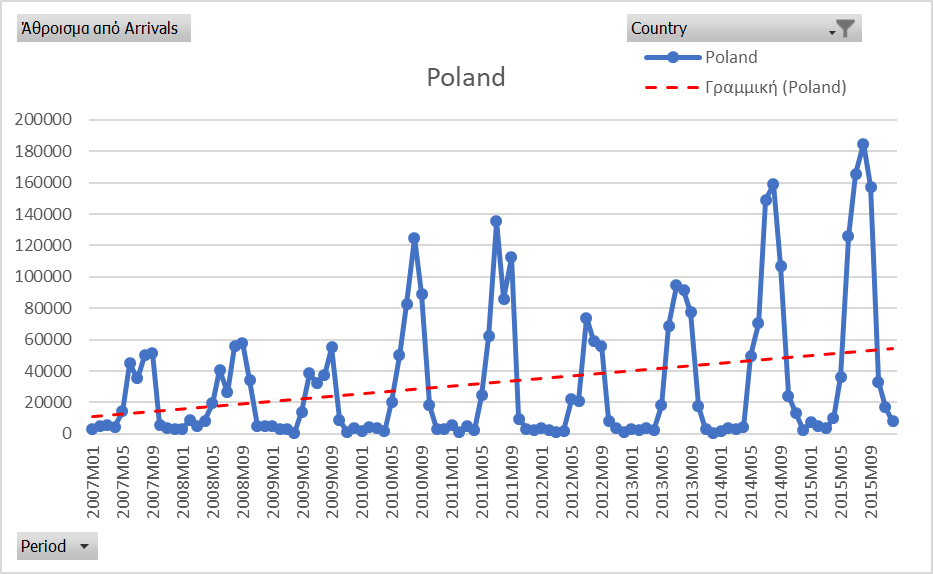


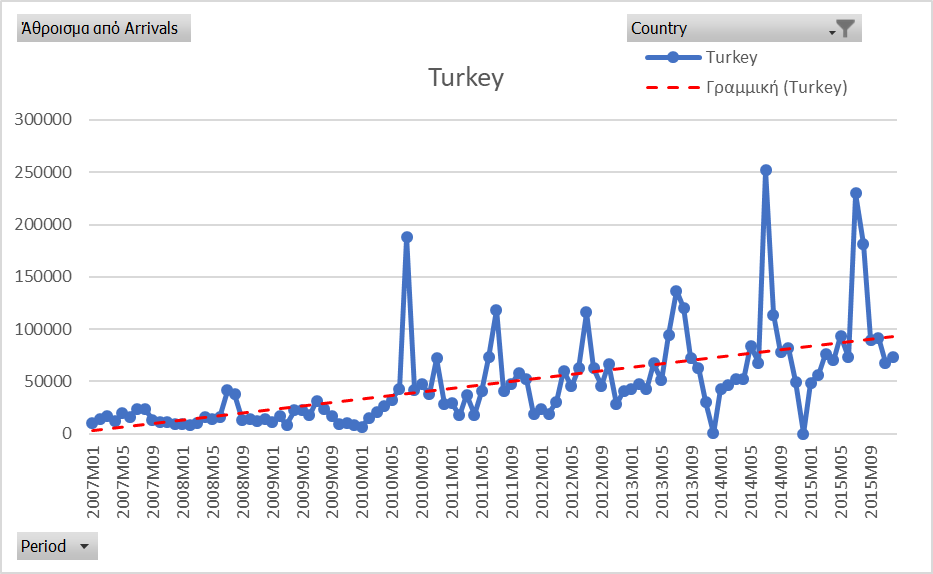


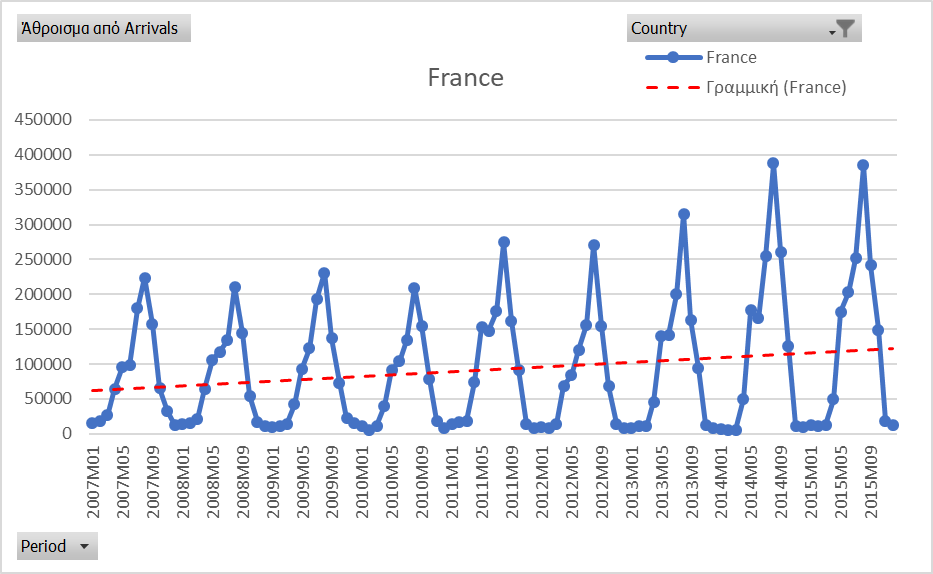


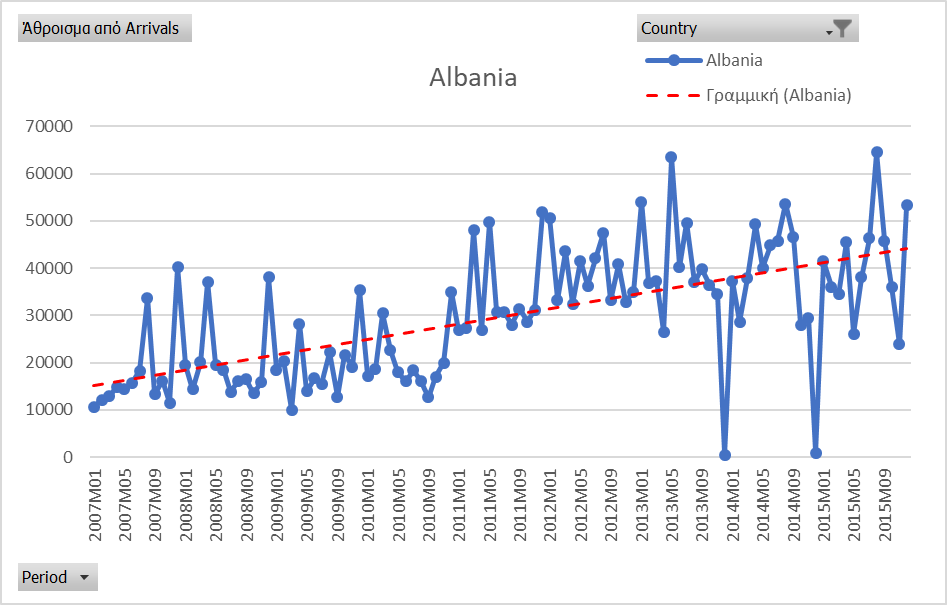
### Countries with positive trend

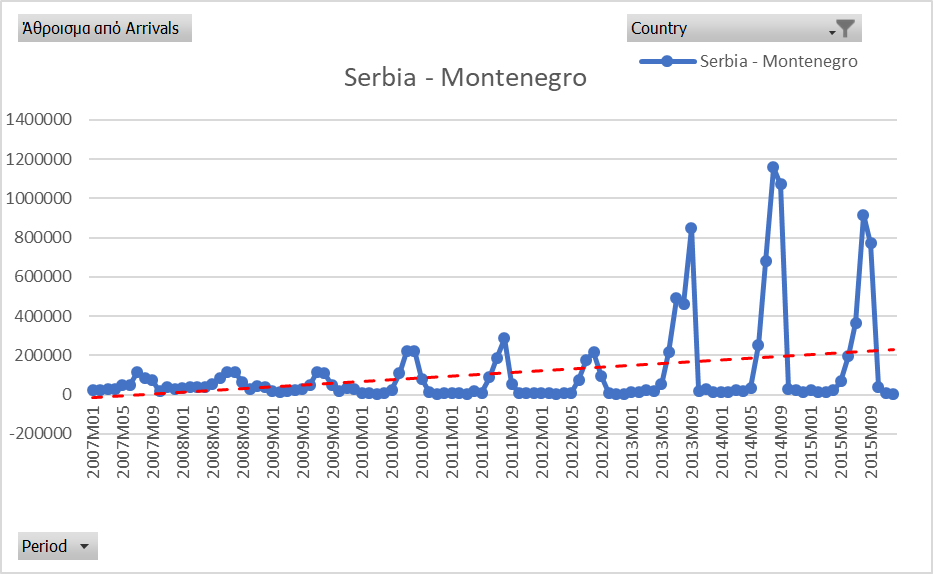


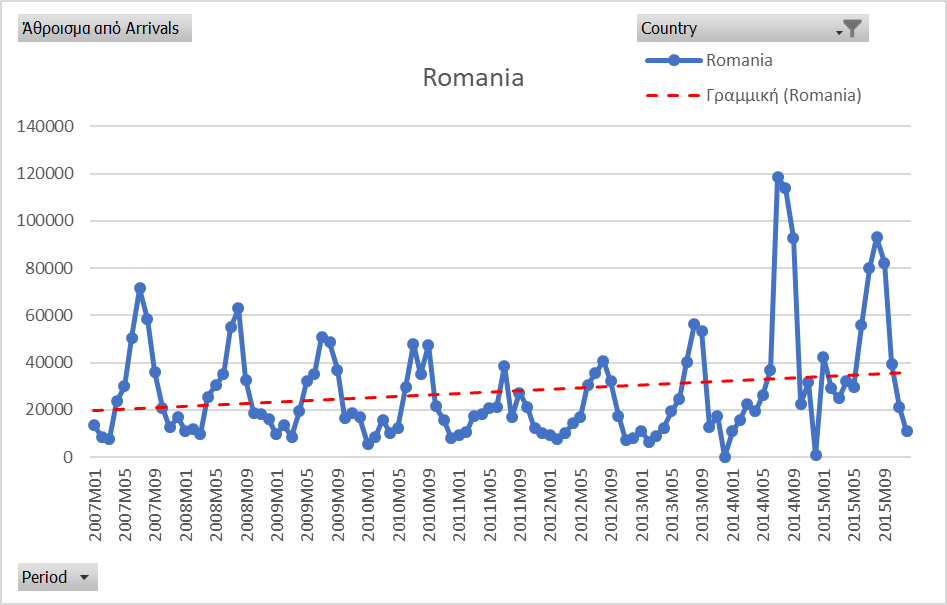


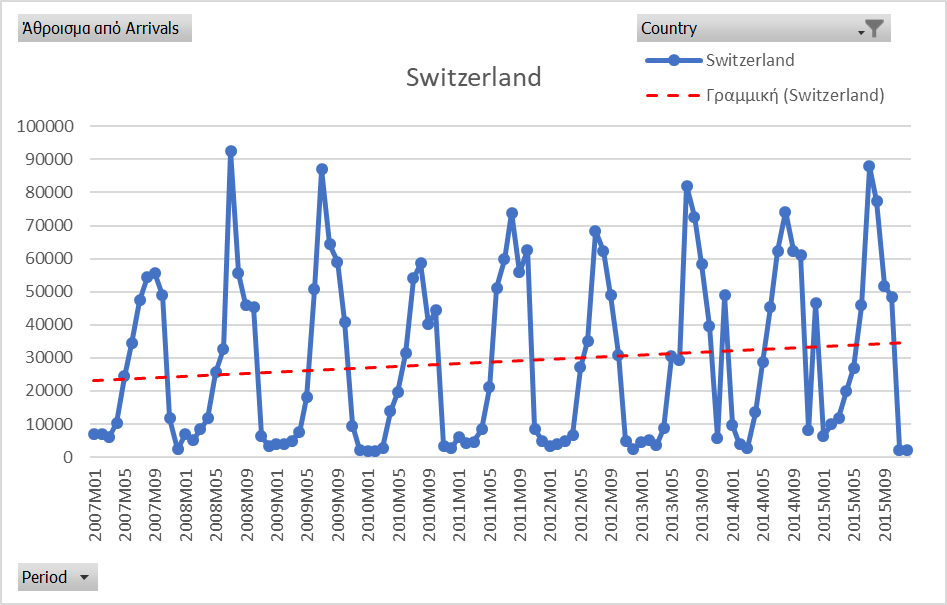


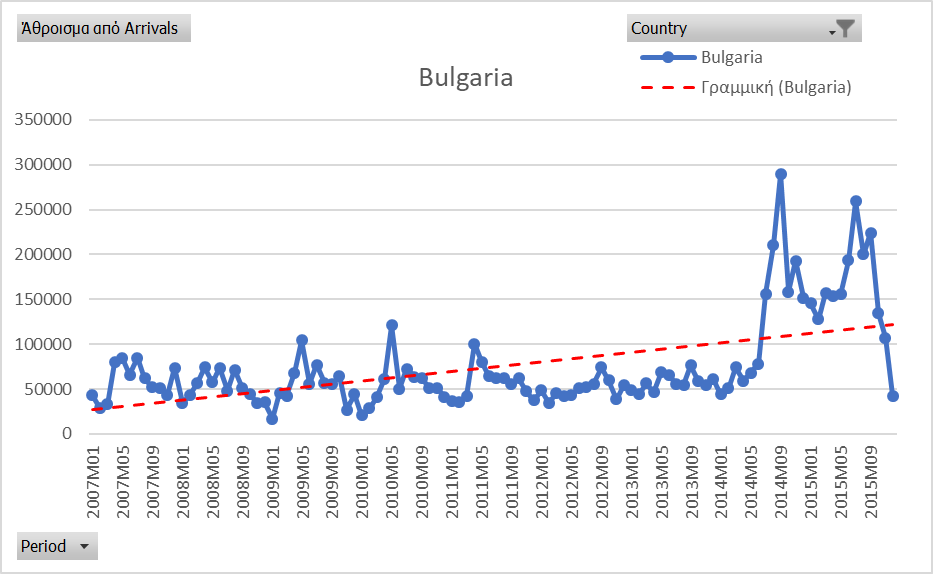


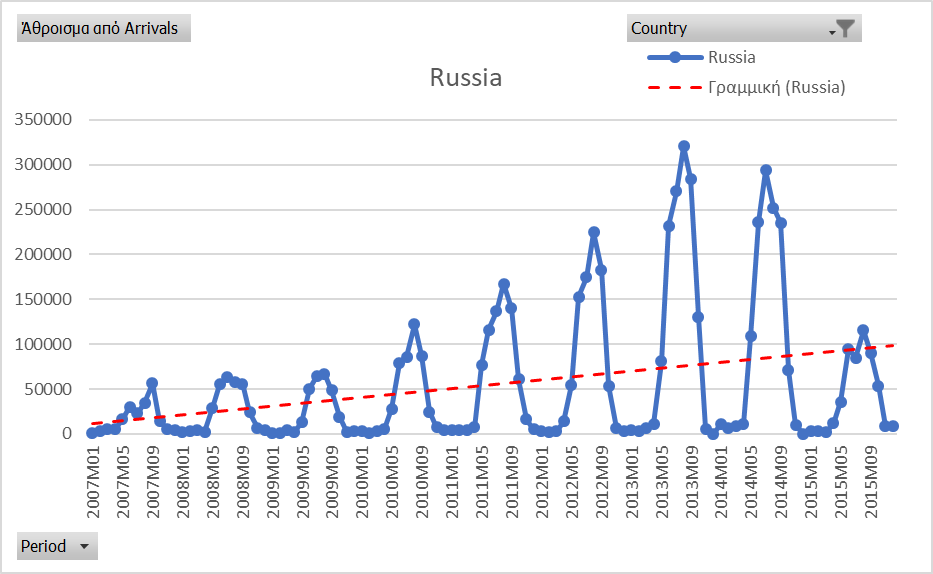




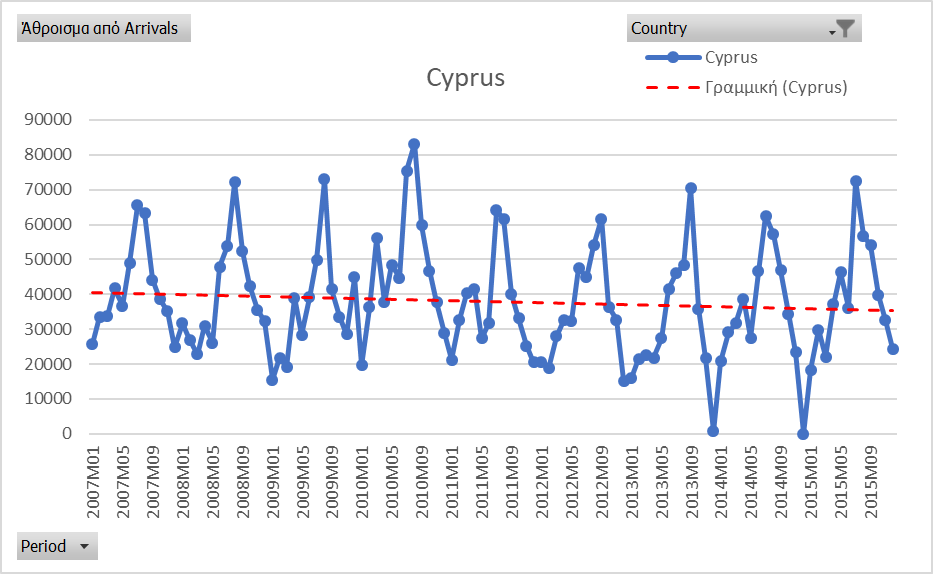








### Country with negative trend



# **Forecasting**

## **3.1. Introduction**

The study aims at predicting the tourism flows in Greece. A set of monthly data from ELSTAT is used for the period 2000 - 2015. The variables under consideration include the number of tourist arrivals (total, residential, non-residential) and the number of overnights total. For this purpose, the study employs two forecasting techniques, Triple Exponential Smoothing and Seasonal Autoregressive Integrated Moving-Average (SARIMA). Each approach provides insights into the future tourism flows from a different angle.

The investigation of the methods involves a comparison of the forecasting performance of each approach in order to determine the most accurate one. This can be achieved using a variety of forecast error measurements. In this context, the outcome of multiple forecasting tests will be evaluated in order to decide upon the most robust technique.

Focusing on the total tourist arrivals, Figures 3.1 – 3.4 present the patterns they follow.

**

Figure 3.1. Time-series plot for total number of arrivals

**

Figure 3.2. Boxplot for arrivals in months.

**

Figure 3.3. Boxplot of Arrivals-Total by season.

**

Figure 3.4. Time-series composite graph for Arrivals-Total by Seasons

Figure 3.1 indicates that tourist arrivals are characterized by seasonality. They follow a standard pattern whereby they reach a peak during the summer months and then they fall. According to Figure 3.2 the most active month in terms of arrivals is August, followed by July. This is hardly surprising, given that the most popular tourist destinations in Greece are the islands. Most tourists come from North European countries where the climate and the geographical characteristics do not favor the development of seaside resorts. Therefore, Mediterranean countries, such as Greece, is an attractive holiday destination for North Europeans.

In this context, Figure 3.3 illustrates the categorization of months into 3 tourist seasons: Low, Medium and High. Low season months involve November, December, January, February, and March and correspond to the lowest points in Figure 3.1. Those troughs hover around 500,000 and do not exhibit any trends. Medium season starts from October and ends on April, while the high season is from May to September. There is a trend in high season months, which is declining from period 1 to 60 and becomes increasing from period 60 to 192, with the exception of the interval (140, 160). The trends during the high season are extrapolated to the entire year and determine the peaks and troughs. Any annual increase or decrease in tourist arrivals comes as a result of equivalent fluctuations during those months. This is also illustrated by Figures 3.2 – 3.4; those boxplots indicate high variability on the high season months.

Another noteworthy observation is that the number of tourist arrivals is relatively low from 2000 to 2004 with a slightly decreasing trend. There was a spike in 2004 and since then they remain at higher levels. This increase is mainly attributed to Athens 2004 Olympics that gave a boost to Greek tourism.

## **3.2 Short review on forecasting methods in tourism**

Due to the strong seasonality of the data, the traditional forecasting methods such as simple exponential smoothing, ARIMA etc. produced unacceptable results – the deviations between observed and forecasted values were found to be 20 – 30% - and thus they were discarded.

The Box-Jenkins approach is one of the most widely used methodologies in tourism forecasting and many authors find that it outperforms alternative techniques.

Kim & Uysal (1998) develop an ARMAX model for the numbers of nights spent in hotels in Seoul, Korea, and they find that the variables: price of rooms, trade volume, and events organized, such as conferences and exhibitions, have a significant impact on hotel demand.

Cho (2001) focuses on Hong Kong tourist arrivals and formulates an ARIMAX type model (adjusted ARIMA). This model involves ARIMA residuals of the following variables: GDP/GNP, Consumer Price Index (CPI), money supply, unemployment rate, imports, exports, and discount rate. The results suggest that ARIMAX is the most accurate forecasting method for inbound tourists from Japan, while univariate ARIMA is the best predictor for tourists coming from the US and UK. Interestingly, according to the results, both of these models appear to perform better than exponential smoothing.

Akal (2004) attempts to predict the tourism revenues in Turkey using tourist arrivals as an explanatory variable in an AR(I)MAX model. Then the author shows that the specified model outperforms the simple cause–effect technique.

Gounopoulos, Petmezas, & Santamaria (2012) provide evidence that ARIMA performs better than exponential smoothing models. ShuiKi, ShinHuei, & ChiKeung (2013) and Baldigara and Mamula (2015) view SARIMA type models as the most appropriate framework to predict tourist arrivals and tourist flows respectively. In particular, ShuiKi et al. (2013) compare SARIMA with Seasonal moving average and Holt-Winter models and find that SARIMA produces more accurate forecasts. More recently, Hassani, Silva, Antonakakis, Filis, & Gupta (2017) attempt to predict tourist arrivals to Europe using ARIMA among other parametric and non-approaches, such as Autoregressive Fractionally Integrated Moving Average (ARFIMA), Exponential Smoothing (ETS), Neural Networks (NN), Trigonometric Box-cox ARMA Trend Seasonal model (TBATS), Moving Average (MA), Weighted Moving Average (WMA), Singular Spectrum Analysis (SSA-R and SSA-V). The comparison reveals that none of these models stands out.

A more detailed analysis of the relevant literature is provided in Appendix IV.

Two seasonality-based methods have been selected as the most appropriate for the Greek tourism data: Triple exponential smoothing, the Holt-Winters method, and Seasonal Autoregressive Integrated Moving-Average (SARIMA). Triple exponential smoothing breaks forecasts into three components, i.e. level, trend and seasonality, and is suitable for times series that exhibit trend and seasonality, such as the ones analyzed in this study. SARIMA extends the typical ARIMA model by including an additional seasonal term.

## **3.3. Methods**

### **3.3.1. Exponential Smoothing**

Exponential Smoothing is a widely used forecasting technique. It is based on the most recent forecast, the actual value for the previous period, and one or more smoothing parameters. The distinctive characteristic of this method is that recent observations are assigned more weight than older ones.

Depending on the number of smoothing constants that are incorporated into the model, there are three variations of Exponential Smoothing:

* Single Exponential Smoothing which is the simplest form and uses only one parameter, the smoothing constant *a* that corresponds to the reaction rate to differences between forecast and actual values.
* Exponential Smoothing with Trend Adjustment (Holt’s Model), which incorporates an additional parameter, the smoothing constant β. This parameter is essentially an adjustment for trend.
* Exponential Smoothing with Trend and Seasonality or Triple Exponential Smoothing, the Holt-Winters (HW) Method.

This study will consider the third and more complex form of Exponential Smoothing, given that the data show seasonality, besides trend.

The equations to compute the forecasts are as follows:







where Ft the exponentially smoothed forecast without trend for period t, FITt the forecast including trend, and Tt the exponentially smoothed trend for period t.

The Triple Exponential Smoothing includes one more constant, γ, which is the smoothing parameters for the seasonal component.

The above mentioned smoothing constants determine the sensitivity of forecasts to changes. When the parameters are given high values, they become more responsive to recent levels, rather than older estimates.

### **3.3.2. The Box-Jenkins approach (ARMA) – SARIMA framework**

Box-Jenkins (1970) introduced the AutoRegressive Moving Average (ARMA) model. This model encompasses the AR(p) and MA(q) models. An ARMA(p,q) model can be mathematically represented as follows:



where Xt is a time-series c is a constant, the random variables εt, εt-i are white noise, p the order of autocorrelation, q the order of the moving average, and φi, θi are parameters.

An important limitation of the ARMA models is that they require stationary data. This inevitably narrows down their range of applications, since most series have unit roots. Yet, Box and Jenkins managed to overcome this weakness within the ARMA framework. They handled this by differencing the successive observations of a series as many times as it is necessary until they become stationary.

In this context, non-stationary time-series can be modelled using an extension of ARMA which is known as AutoRegressive Integrated Moving Average ARIMA (p,d,q). ARIMA (p,d,q) is based on ARMA (p,q) and includes an extra parameter d representing the number of differencing to achieve stationarity.

Another modification of ARMA is the Seasonal ARIMA (SARIMA) model, which can be employed when seasonality is present. The SARIMA model can be expressed as SARIMA (p,d,q) (P,D,Q), where the capital letters P,D,Q are the counterparts of the parameters p,d,q for the seasonal model.

The ARIMA (or 'Box-Jenkins approach) involves the following modelling steps:

1) Identification

Initially, the outcome of stationarity tests dictates if the modelling framework should be based on ARMA or ARIMA. Also, correlograms are used to specify the appropriate parameters (p,q).

2) Model Specification

The models specification can be broken down into the following sub-steps:

2.1) The most reliable criterion would be based on the precision of results (Yokum and Armstrong, 1995). Given that there is no such measure which is universally accepted, the model is specified on the basis of the examination of the autocorrelation (ACF) and partial autocorrelation (PACF) functions.

2.2) The optimal lag length can also be determined on the basis of the Akaike Information Criterion (AIC) or the Schwarz criterion.

3) Estimation

Based on the values (p, q), the ARMA terms are estimated using Least Squares.

4) Diagnostic checking

This step includes residual diagnostics (e.g. residual serial correlation and residual heteroscedasticity tests)

5)Forecasting

Generation of ex-post and ex-ante forecasts.

6) Evaluation of the forecasting accuracy

The performance of the model is assessed on the basis of the following criteria: Mean Absolute Deviation (MAD), Mean absolute percent error (MAPE), Mean Absolute Deviation Percent (MADP), and Mean squared deviation (MSD)

The SARIMA model

The Seasonal ARIMA (SARIMA) model, is an extension of ARIMA (p,d,q) whereby a seasonal factor is added. The SARIMA model can be written asSARIMA (p,d,q) (P,D,Q), where the capital letters P,D,Q refer to the counterparts of the parameters p,d,q for the seasonal model.

## **3.4. Model fitting**

For the Triple exponential smoothing, the multiplicative method has been preferred as the seasonal data variations are not approximately constant but they are changing proportionally through the series. The fitting procedure involved the estimation of the values for the parameters α, β and γ (α=0,15, β=0,06 γ=0,92) so to get error values as minimum as possible. This has been achieved first by selecting greater smoothing effect on level and trend (low values for α and β) and limited smoothing for seasonality (high values for γ) and second by applying a trial and error procedure on a great number of different value combinations.

As far as SARIMA modelling is concerned, the most suitable model proved to be a SARIMA (4,1,2) (1,1,1) for tourist arrivals, a SARMA (3,4) (1,1) for nights total and arrivals – residents, and a SARIMA (3,1,3) (1,1,1) for arrivals non-residents. Seasonality was present in all time-series, therefore we selected SAR(I)MA, rather than AR(I)MA models.

The model selection was based initially on visual inspection of the graph, as well as on stationarity tests (ADF, KPSS) in order to determine the level of differencing. The tourist arrivals were found non-stationary in levels, therefore we applied first differencing to account for the first order of integration. According to the relevant tests, the nights arrivals and the arrivals – residents are stationary, therefore they were both modelled using a SARMA model. Finally, the arrivals-non-residents were found non-stationary and they were handled in a SARIMA framework.

It should be noted that the variable ‘arrivals-residents’ was the only one that was log-transformed, in order to avoid the presence of heteroscedasticity as the change in the growth rate varied over time.

The parameters of the SAR(I)MA models were determined using the Akaike Information Criterion (AIC) in combination with an examination of the autocorrelation (ACF) and partial autocorrelation (PACF) functions.

## **3.5. Evaluation of forecasting accuracy**

The evaluation of the accuracy of model predictions has always been a challenging endeavor and in fact determines the success or failure of each modelling approach. Therefore, it is important to use sound evaluation criteria and ensure the validity of results.

Forecasting accuracy is judged by means of forecast errors, which are based on the difference between the forecast value and the actual value of each observation. Error occurs even in the best forecasts, but in varying degrees. Error may come from pure randomness of from the bias caused by a consistent mistake.

This study adopts a variety of forecast errors. Specifically, the following criteria will be used to assess the predictive power of each model: Mean Absolute Deviation (MAD), Mean absolute percent error (MAPE), Mean Absolute Deviation Percent (MADP), and Mean squared deviation (MSD)

Those statistics constitute relative measures that can be used for comparison purposes of the same series. The smaller the error, the greater the forecasting accuracy of a given model.

A description of the forecast error statistics used in this study is presented after a brief description of R squared.

* **R squared**

The coefficient of determination (R2) measures how well the data fit the model. It generally describes the goodness of fit of the regression line to the data.

TSS stands for the Total Sum of Squares and is given by:



RSS is the Residual Sum of Squares (RSS):



where Y denotes the independent variable,  the sample mean and ut the residual.

The value of R-squared is defined as:



* **Mean Absolute Error (MAE)**

The mean absolute error is defined as:



where Ft is the forecast value and At is the actual value.

The MAE represents the average of absolute errors, indicating the proximity of predictions to the actual values. Therefore, the larger the MAE the less accurate the model.

* **Mean absolute percent error (MAPE)**

In the same notation, the mean absolute percent error is defined as:



MAPE remedies a major weakness of MAE, which is, its dependence on the scaling of the variable. This dependence can become a serious concern when the model comprises varying time intervals or different variables. The way MAPE overcomes this issue is by the division of the difference (Ft – At) by the actual value At.

However, this technique presumes that there will be no zero values in the sample; otherwise the use of this metric will become problematic, since the division with zero is not possible

* **Mean Absolute Deviation Percent (MADP)**

MADP derives from the division of MAD by the average of the actuals.

This forecast measure has remedies one of the weaknesses of MAPE, as it does not skew the error rates that approach zero.

The following formula is used for the computation of MADP:



* **Mean squared deviation (MSD)**

MSD is an alternative forecast error measure which is given by the following formula:



where n is the number of forecasts.

However, the major weakness of this measure is that outliers tend to have a greater impact on MSD than on MAD.

## **3.6. Application and Results**

The evaluation of the two proposed methods was done by means of ex-ante forecasting analysis whereby five different samples of the dataset were specified, with the remaining observations to be considered unknown. The resized samples were used to produce out-of-sample forecasts and ultimately compare the forecasted values with the actual ones.

Then, we conduct out-of-sample analysis for the monthly values of 2016-2018. Due to the lack of data for this period, it is not possible to compare the forecasts with actual values. However, the monthly forecasts are compared with the respective observations of 2015 and we calculate the average rate per year.

For each of the tests, the observed and forecasted values were used to estimate the typical error formulas MAPE, MAD, MADP, MSD and R squared (see spreadsheet).

### **3.6.1. Tourist Arrivals**

For tourist arrivals, the first sub-sample spanned 2000-2010 with the aim to forecast each month of 2011, the second started from 2000 to 2011 to forecast the monthly values of 2012 etc. We follow the same procedure for each year, up to 2015, when the available data end. Then we conduct out-sample analysis for 2016-2018.

Below are the forecasting tests for each sub-sample.

**TEST 1 : 2000 - 2010 => 2011**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | HW-Exponential Smoothing | SARIMA |
| Period No | Month | Observed | Forecast | Forecast |
| 133 | 2011M01 | 524830 | 532524,5516 | 516283.3531 |
| 134 | 2011M02 | 515934 | 547524,9946 | 577678.6746 |
| 135 | 2011M03 | 689867 | 678809,3093 | 704639.6383 |
| 136 | 2011M04 | 1069357 | 1013687,828 | 1067115.697 |
| 137 | 2011M05 | 1737712 | 1662439,412 | 1727533.762 |
| 138 | 2011M06 | 2299967 | 1972203,227 | 2047092.934 |
| 139 | 2011M07 | 2818308 | 2581278,775 | 2682844.725 |
| 140 | 2011M08 | 2923458 | 2721735,774 | 2850771.721 |
| 141 | 2011M09 | 2156203 | 1975487,505 | 2072624.755 |
| 142 | 2011M10 | 1078490 | 1088554,312 | 1156130.385 |
| 143 | 2011M11 | 466377 | 525427,7895 | 573136.5152 |
| 144 | 2011M12 | 465103 | 522826,1801 | 565965.6897 |
| Errors : |  |  |  |  |
| MAPE |  |  | **6.89%** | **7.54%** |
| MAD |  |  | **104613** | **77279** |
| MADP |  |  | **7.5%** | **5.5%** |
| MSD |  |  | **21152219307** | **10531284698** |
| R2 |  |  | **0.974** | **0.9872** |

|  |  |
| --- | --- |
|  |  |
| HW-Exponential Smoothing | SARIMA |

**TEST 2 : 2000 - 2011 => 2012**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | HW-Exponential Smoothing | SARIMA |
| Period No | Month | Observed | Forecast | Forecast |
| 145 | 2012M01 | 452558 | 529345,8144 | 477904.236 |
| 146 | 2012M02 | 424357 | 526276,1435 | 504744.3777 |
| 147 | 2012M03 | 581144 | 696948,0355 | 701007.1143 |
| 148 | 2012M04 | 858976 | 1068070,753 | 1078114.96 |
| 149 | 2012M05 | 1442049 | 1725995,788 | 1750908.11 |
| 150 | 2012M06 | 2027151 | 2216641,942 | 2274888.063 |
| 151 | 2012M07 | 2500205 | 2704051,44 | 2822165.213 |
| 152 | 2012M08 | 2679376 | 2792031,219 | 2928813.823 |
| 153 | 2012M09 | 1965510 | 2041464,269 | 2177569.568 |
| 154 | 2012M10 | 973555 | 1035984,214 | 1116535.551 |
| 155 | 2012M11 | 445345 | 461308,798 | 497083.2787 |
| 156 | 2012M12 | 440470 | 468135,2699 | 500133.7975 |
| Errors : |  |  |  |  |
| MAPE |  |  | **12.23%** | **14.76%** |
| MAD |  |  | **122963** | **169931** |
| MADP |  |  | **10.0%** | **13.8%** |
| MSD |  |  | **21239908613** | **38649011460** |
| R2 |  |  | **0.9683** | **0.9424** |

|  |  |
| --- | --- |
|  |  |
| HW-Exponential Smoothing | SARIMA |

**TEST 3 : 2000 - 2012 => 2013**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | HW-Exponential Smoothing | SARIMA |
| Period No | Month | Observed | Forecast | Forecast |
| 157 | 2013M01 | 427528 | 426146,2436 | 463300.6838 |
| 158 | 2013M02 | 441053 | 410914,7072 | 400179.7341 |
| 159 | 2013M03 | 620089 | 570310,9054 | 558670.3416 |
| 160 | 2013M04 | 846254 | 861975,6968 | 837081.1257 |
| 161 | 2013M05 | 1724551 | 1460119,937 | 1438784.894 |
| 162 | 2013M06 | 2306252 | 2034483,351 | 2041531.486 |
| 163 | 2013M07 | 2687098 | 2498713,397 | 2508392.256 |
| 164 | 2013M08 | 2942170 | 2646551,441 | 2683293.28 |
| 165 | 2013M09 | 2201812 | 1923836,379 | 1938428.508 |
| 166 | 2013M10 | 1154154 | 950101,6708 | 922852.1238 |
| 167 | 2013M11 | 504584 | 429996,9664 | 398681.4433 |
| 168 | 2013M12 | 470217 | 424267,8947 | 386920.9429 |
| Errors : |  |  |  |  |
| MAPE |  |  | **9.67%** | **11.90%** |
| MAD |  |  | **143316** | **151599** |
| MADP |  |  | **10.53%** | **11.10%** |
| MSD |  |  | **33073223173** | **33210318761** |
| R2 |  |  | **0.9607** | **0.9605** |

|  |  |
| --- | --- |
|  |  |
| HW-Exponential Smoothing | SARIMA |

**TEST 4 : 2000 - 2013 => 2014**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | HW-Exponential Smoothing | SARIMA |
| Period No | Month | Observed | Forecast | Forecast |
| 169 | 2014M01 | 479484 | 466341,4965 | 430056.4317 |
| 170 | 2014M02 | 480858 | 475990,2277 | 432114.8437 |
| 171 | 2014M03 | 647404 | 664440,3762 | 618373.1828 |
| 172 | 2014M04 | 1034259 | 922710,1202 | 838539.3112 |
| 173 | 2014M05 | 1989444 | 1820893,505 | 1674549.314 |
| 174 | 2014M06 | 2450010 | 2420732,804 | 2239565.885 |
| 175 | 2014M07 | 2902722 | 2832941,063 | 2645653.58 |
| 176 | 2014M08 | 3117313 | 3082170,866 | 2897020.846 |
| 177 | 2014M09 | 2302398 | 2285592,497 | 2179409.916 |
| 178 | 2014M10 | 1300758 | 1175791,681 | 1155433.396 |
| 179 | 2014M11 | 522399 | 511908,5397 | 526716.9188 |
| 180 | 2014M12 | 516445 | 480507,3175 | 498145.9965 |
| Errors : |  |  |  |  |
| MAPE |  |  | **4.14%** | **8.76%** |
| MAD |  |  | **53129** | **134713** |
| MADP |  |  | **3.59%** | **9.1%** |
| MSD |  |  | **5466763011** | **28218747213** |
| R2 |  |  | **0.9942** | **0.9702** |

|  |  |
| --- | --- |
|  |  |
| HW-Exponential Smoothing | SARIMA |

**TEST 5 : 2000 - 2014 => 2015**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | HW-Exponential Smoothing | SARIMA |
| Period No | Month | Observed | Forecast | Forecast |
| 181 | 2015M01 | 485431 | 508496,3522 | 521850.2273 |
| 182 | 2015M02 | 532593 | 510920,5711 | 517061.629 |
| 183 | 2015M03 | 689505 | 693725,302 | 713854.9207 |
| 184 | 2015M04 | 1141786 | 1079975,832 | 1081673.757 |
| 185 | 2015M05 | 2152897 | 2062297,636 | 2074412.458 |
| 186 | 2015M06 | 2573973 | 2562402,387 | 2549569.955 |
| 187 | 2015M07 | 3025144 | 3036181,564 | 2996084.426 |
| 188 | 2015M08 | 3247472 | 3273699,86 | 3211211.302 |
| 189 | 2015M09 | 2493173 | 2426855,878 | 2396866.454 |
| 190 | 2015M10 | 1351422 | 1347418,242 | 1373261.707 |
| 191 | 2015M11 | 559050 | 546044,4768 | 602945.5154 |
| 192 | 2015M12 | 568733 | 535136,2409 | 588124.9662 |
| Errors : |  |  |  |  |
| MAPE |  |  | **2.66%** | **3.55%** |
| MAD |  |  | **30594** | **40505** |
| MADP |  |  | **1.95%** | **2.58%** |
| MSD |  |  | **1641978459** | **2228616768** |
| R2 |  |  | **0.9984** | **0.9978** |

|  |  |
| --- | --- |
|  |  |
| HW-Exponential Smoothing | SARIMA |

**TEST 6 : 2000 - 2015 => 2016, 2017, 2018**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | HW-Exponential Smoothing | | SARMA | |
| Month | Observed 2015 | Forecast | % Difference | Forecast | % Difference |
| 2016-01 | 485431 | 520177 | 7,16% | 531431 | 9.48% |
| 2016-02 | 532593 | 563375 | 5,78% | 582689 | 9.41% |
| 2016-03 | 689505 | 733203 | 6,34% | 748670 | 8.58% |
| 2016-04 | 1141786 | 1199478 | 5,05% | 1199703 | 5.07% |
| 2016-05 | 2152897 | 2253558 | 4,68% | 2197370 | 2.07% |
| 2016-06 | 2573973 | 2706352 | 5,14% | 2602449 | 1.11% |
| 2016-07 | 3025144 | 3190640 | 5,47% | 3073075 | 1.58% |
| 2016-08 | 3247472 | 3436208 | 5,81% | 3288380 | 1.26% |
| 2016-09 | 2493173 | 2624940 | 5,29% | 2538843 | 1.83% |
| 2016-10 | 1351422 | 1427648 | 5,64% | 1422885 | 5.29% |
| 2016-11 | 559050 | 588749 | 5,31% | 632527 | 13.14% |
| 2016-12 | 568733 | 593321 | 4,32% | 645022 | 13.41% |
| Average 2016 | |  | **5,50%** |  | **6.02%** |
| 2017-01 | 485431 | 544293 | 12,13% | 609156 | 25.49% |
| 2017-02 | 532593 | 589394 | 10,66% | 665537 | 24.96% |
| 2017-03 | 689505 | 766936 | 11,23% | 833999 | 20.96% |
| 2017-04 | 1141786 | 1254452 | 9,87% | 1285166 | 12.56% |
| 2017-05 | 2152897 | 2356448 | 9,45% | 2301412 | 6.90% |
| 2017-06 | 2573973 | 2829447 | 9,93% | 2719439 | 5.65% |
| 2017-07 | 3025144 | 3335214 | 10,25% | 3178204 | 5.06% |
| 2017-08 | 3247472 | 3591323 | 10,59% | 3398995 | 4.67% |
| 2017-09 | 2493173 | 2742989 | 10,02% | 2648944 | 6.25% |
| 2017-10 | 1351422 | 1491613 | 10,37% | 1510489 | 11.77% |
| 2017-11 | 559050 | 615030 | 10,01% | 721129 | 28.99% |
| 2017-12 | 568733 | 619707 | 8,96% | 731224 | 28.57% |
| Average 2016 | |  | **10,29%** |  | **15.15%** |
| 2018-01 | 485431 | 568410 | 17,09% | 696291 | 43.44% |
| 2018-02 | 532593 | 615413 | 15,55% | 748515 | 40.54% |
| 2018-03 | 689505 | 800668 | 16,12% | 916634 | 32.94% |
| 2018-04 | 1141786 | 1309425 | 14,68% | 1368129 | 19.82% |
| 2018-05 | 2152897 | 2459337 | 14,23% | 2369900 | 10.08% |
| 2018-06 | 2573973 | 2952541 | 14,71% | 2777349 | 7.90% |
| 2018-07 | 3025144 | 3479787 | 15,03% | 3247712 | 7.36% |
| 2018-08 | 3247472 | 3746438 | 15,36% | 3464465 | 6.68% |
| 2018-09 | 2493173 | 2861038 | 14,75% | 2716109 | 8.94% |
| 2018-10 | 1351422 | 1555578 | 15,11% | 1597822 | 18.23% |
| 2018-11 | 559050 | 641310 | 14,71% | 808772 | 44.67% |
| 2018-12 | 568733 | 646093 | 13,60% | 821765 | 44.49% |
| Average 2016 | |  | **15,08%** |  | **23.76%** |

First of all, it should be noted that R-squared is very high, as it is above 0.96 in all cases and many times it exceeds 0.99. This is indicative of the goodness of fit of the proposed models.

Overall, the HW approach appears to be superior to SARIMA (2,1,4) (1,1,1). All of the forecasting exercises indicate that the forecast errors of HW are lower than those of SARIMA. The difference is not too wide, as it ranges between 0.65% (Test 1) to 4.62% (Test 4) in terms of MAPE. Notably in Test 5, the HW reaches a high level of accuracy (MAPE: 2.66%, MADP: 1.95%). On the other hand, SARIMA achieves its best performance again in Test 5 (MAPE 3.55%, MADP 2.58%), but its accuracy is still lower than HW’s. The fact that both models are more effective in Test 5 is mainly attributed to the largest number of observations. The sample size for this test is larger than for Tests 1 – 4 and this enables the models to track the underlying patterns more effectively. Furthermore, the R2 obtained in Test 5 is the highest for both models. This suggests that when the dataset is updated with new observations, there is upside potential for both models in terms of their performance. However, it should be noted that SARIMA models are more dependent on sample size than Exponential Smoothing, which relies more heavily on the prior period of each forecast. Yet, a larger sample enhances also the adaptive forecasting employed under the HW approach as it facilitates the better selection of β and γ smoothing constants.

A point that deserves special attention is the way the two methods behave in the tourist arrivals’ fluctuations from 2011 to 2013. Focusing on August’s peak, it is observed that arrivals dropped from 2,923,458 in 2011 to 2,679,376 in 2012 and they went back up again in August 2013 (2,942,170). It turns out that none of the models manages to track those changes. Both models predict an increase for 2012 and a decrease in 2013 (with HW generating forecasts which are closer to the actual values). This failure has its roots in the structure of those models. They both rely on past values and they largely base their predictions on patterns of past periods. In all of the above-mentioned cases, the forecasts are affected by the trend of the previous year and this prevents them from capturing those unexpected changes. This goes down as a weakness of this kind of modelling approaches. However, we should also take into account the small number of observations in those tests. Especially ARIMA models require large data sets in order to perform satisfactorily.

For 2016-2018 both models predict a rising trend of tourist arrivals, which is in line with expectations. With 2015 as a reference year, Exponential Smoothing estimates a 5.5% average increase for 2016, 10.3% for 2017, and 15.1% for 2018. The corresponding figures for SARIMA are 6%, 15.2%, and 23.8% respectively. It seems that SARIMA provides more optimistic projections for the long term, drawing on the rising trend that came up from 2014 onwards. It should be noted that the growing number of arrivals is not limited to the high season, but they also spread to the middle and low seasons.

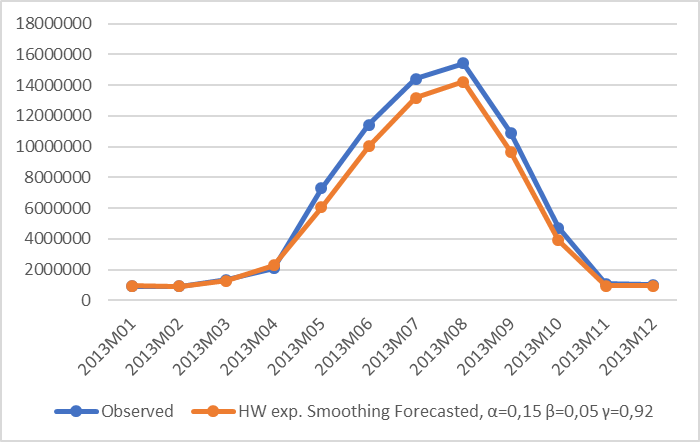
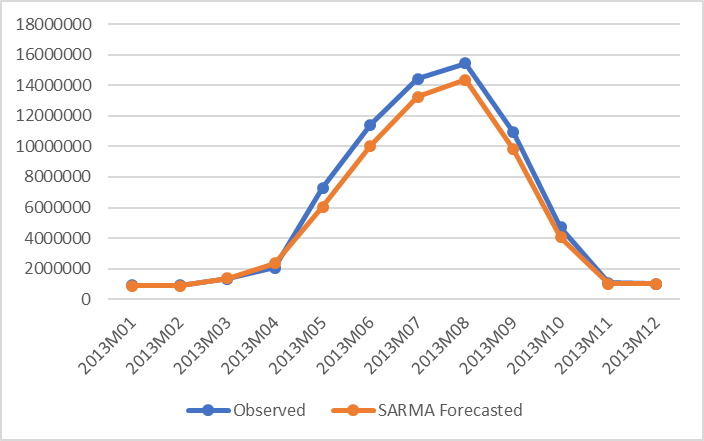
### **3.6.2. Nights – Total**

Nights total are divided into 4 sub-samples: The first one is from 2000 to 2012 and the forecast period is 2013, the second is from 2000 to 2013 with the purpose of forecasting the monthly values of 2014, the third one is from 2000 to 2014, with the forecast period being 2015, and finally, as before, we conduct out-of-sample analysis for 2016 – 2018.

Below are the forecasting tests for each sub-sample.

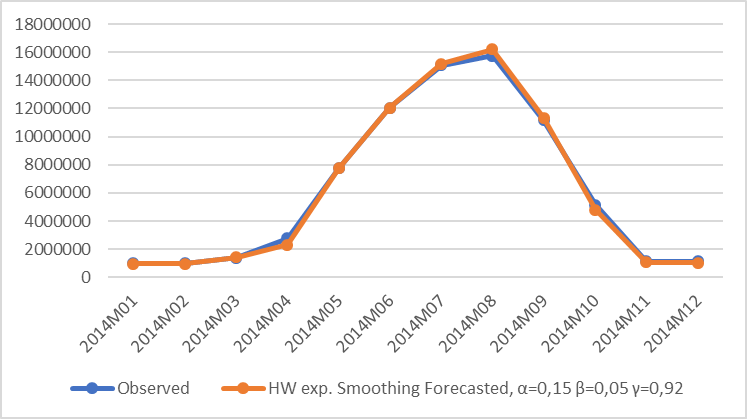
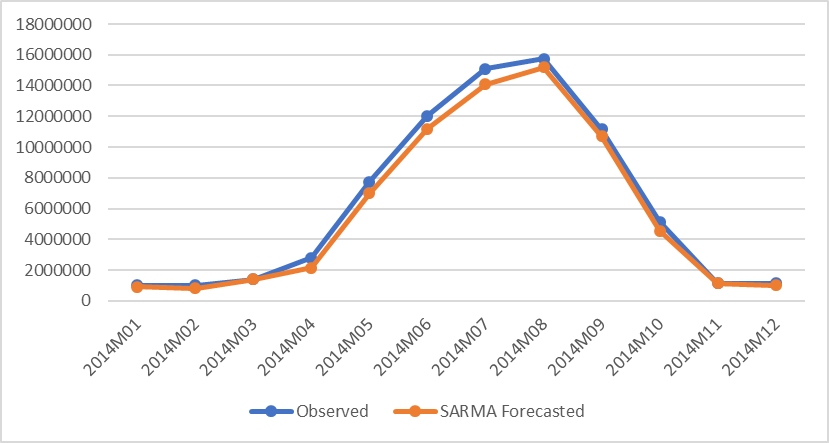
**TEST 1: 2000 - 2012 => 2013**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Period No | Month | Observed | HW Forecast | SARMA Forecast |
| 157 | 2013M01 | 904955 | 936116 | 889961 |
| 158 | 2013M02 | 905347 | 909640 | 889255 |
| 159 | 2013M03 | 1327380 | 1293263 | 1368634 |
| 160 | 2013M04 | 2075559 | 2294957 | 2345894 |
| 161 | 2013M05 | 7280843 | 6058044 | 6042948 |
| 162 | 2013M06 | 11411713 | 10034667 | 10037858 |
| 163 | 2013M07 | 14418211 | 13182747 | 13260248 |
| 164 | 2013M08 | 15422707 | 14225147 | 14350920 |
| 165 | 2013M09 | 10907114 | 9656625 | 9813667 |
| 166 | 2013M10 | 4720176 | 3901027 | 4039961 |
| 167 | 2013M11 | 1066638 | 952261 | 1019830 |
| 168 | 2013M12 | 1028546 | 961903 | 1000943 |
| **Errors :** |  |  |  |  |
| **MAPE** |  |  | **9.02%** | **7.92%** |
| **MAD** |  |  | **631041** | **586021** |
| **MADP** |  |  | **10.6%** | **0.0000020%** |
| **MSD** |  |  | **721213006270** | **637167106892** |
| **R2** |  |  | **0.9756** | **0.9784** |

**TEST 2: 2000 - 2013 => 2014**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Period No | Month | Observed | HW Forecast | SARMA Forecast |
| 169 | 2014M01 | 1011944 | 984737 | 927393 |
| 170 | 2014M02 | 1017317 | 986746 | 804270 |
| 171 | 2014M03 | 1386833 | 1442674 | 1387425 |
| 172 | 2014M04 | 2789288 | 2318459 | 2152589 |
| 173 | 2014M05 | 7753192 | 7755269 | 6977704 |
| 174 | 2014M06 | 12040088 | 12042344 | 11198975 |
| 175 | 2014M07 | 15084228 | 15178375 | 14085440 |
| 176 | 2014M08 | 15731420 | 16198844 | 15200416 |
| 177 | 2014M09 | 11148789 | 11328773 | 10692549 |
| 178 | 2014M10 | 5149045 | 4796008 | 4547384 |
| 179 | 2014M11 | 1136678 | 1086957 | 1125866 |
| 180 | 2014M12 | 1141623 | 1056673 | 1012304 |
| **Errors :** |  |  |  |  |
| **MAPE** |  |  | **4.21%** | **8.93%** |
| **MAD** |  |  | **151504** | **439943** |
| **MADP** |  |  | **2.4%** | **0.0000014%** |
| **MSD** |  |  | **51712427970** | **302775986075** |
| **R2** |  |  | **0.9983** | **0.9902** |

**TEST 3: 2000 - 2014 => 2015**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Period No | Month | Observed | HW Forecast | SARMA Forecast |
| 181 | 2015M01 | 1054362 | 1066867 | 1105566 |
| 182 | 2015M02 | 1117262 | 1069349 | 967976 |
| 183 | 2015M03 | 1456316 | 1476722 | 1426479 |
| 184 | 2015M04 | 2994177 | 2843791 | 2731084 |
| 185 | 2015M05 | 8331185 | 8050607 | 7584523 |
| 186 | 2015M06 | 12148665 | 12551711 | 11916078 |
| 187 | 2015M07 | 15419780 | 15804359 | 14861569 |
| 188 | 2015M08 | 16527563 | 16634258 | 15606194 |
| 189 | 2015M09 | 11578112 | 11833333 | 11061044 |
| 190 | 2015M10 | 5288263 | 5386078 | 5042859 |
| 191 | 2015M11 | 1187823 | 1187948 | 1168206 |
| 192 | 2015M12 | 1228834 | 1181679 | 1134566 |
| **Errors :** |  |  |  |  |
| **MAPE** |  |  | **2.47%** | **5.63%** |
| **MAD** |  |  | **150535** | **319050** |
| **MADP** |  |  | **2.3%** | **0.0%** |
| **MSD** |  |  | **41905715854** | **183665841485** |
| **R2** |  |  | **0.9987** | **0.9999** |

**TEST 4: 2000 - 2015 => 2016, 2017, 2018**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **HW** | | | **SARMA** | |
| **Month** | Observed 2015 | Forecast | % Difference | Forecast | | % Difference |
| 2016-01 | 1054362 | 1100623 | 4.39% | 1252185 | | 18.76% |
| 2016-02 | 1117262 | 1153746 | 3.27% | 1248666 | | 11.76% |
| 2016-03 | 1456316 | 1514337 | 3.98% | 1579170 | | 8.44% |
| 2016-04 | 2994177 | 3075617 | 2.72% | 3097068 | | 3.44% |
| 2016-05 | 8331185 | 8536902 | 2.47% | 8187216 | | -1.73% |
| 2016-06 | 12148665 | 12588671 | 3.62% | 12176911 | | 0.23% |
| 2016-07 | 15419780 | 16033611 | 3.98% | 15280738 | | -0.90% |
| 2016-08 | 16527563 | 17181081 | 3.95% | 16357939 | | -1.03% |
| 2016-09 | 11578112 | 12089276 | 4.41% | 11498398 | | -0.69% |
| 2016-10 | 5288263 | 5533209 | 4.63% | 5231903 | | -1.07% |
| 2016-11 | 1187823 | 1240512 | 4.44% | 1234802 | | 3.96% |
| 2016-12 | 1228834 | 1271847 | 3.50% | 1217886 | | -0.89% |
| **Average 2016** | | **3.78%** |  |  | | **3.36%** |
| 2017-01 | 1054362 | 1142307 | 8.34% | 1252519 | | 18.79% |
| 2017-02 | 1117262 | 1197304 | 7.16% | 1253792 | | 12.22% |
| 2017-03 | 1456316 | 1571329 | 7.90% | 1569492 | | 7.77% |
| 2017-04 | 2994177 | 3191006 | 6.57% | 3100645 | | 3.56% |
| 2017-05 | 8331185 | 8856186 | 6.30% | 8171014 | | -1.92% |
| 2017-06 | 12148665 | 13058030 | 7.49% | 12139303 | | -0.08% |
| 2017-07 | 15419780 | 16629560 | 7.85% | 15242468 | | -1.15% |
| 2017-08 | 16527563 | 17817708 | 7.81% | 16305419 | | -1.34% |
| 2017-09 | 11578112 | 12535853 | 8.27% | 11474874 | | -0.89% |
| 2017-10 | 5288263 | 5736977 | 8.49% | 5233499 | | -1.04% |
| 2017-11 | 1187823 | 1286057 | 8.27% | 1252384 | | 5.44% |
| 2017-12 | 1228834 | 1318399 | 7.29% | 1239851 | | 0.90% |
| **Average 2017** | | **7.64%** |  |  | | **3.52%** |
| 2018-01 | 1054362 | 1183990 | 12.29% | 1269817 | | 20.43% |
| 2018-02 | 1117262 | 1240862 | 11.06% | 1274773 | | 14.10% |
| 2018-03 | 1456316 | 1628321 | 11.81% | 1587388 | | 9.00% |
| 2018-04 | 2994177 | 3306395 | 10.43% | 3111848 | | 3.93% |
| 2018-05 | 8331185 | 9175469 | 10.13% | 8161665 | | -2.03% |
| 2018-06 | 12148665 | 13527389 | 11.35% | 12110857 | | -0.31% |
| 2018-07 | 15419780 | 17225509 | 11.71% | 15202283 | | -1.41% |
| 2018-08 | 16527563 | 18454335 | 11.66% | 16259604 | | -1.62% |
| 2018-09 | 11578112 | 12982429 | 12.13% | 11450232 | | -1.10% |
| 2018-10 | 5288263 | 5940746 | 12.34% | 5236315 | | -0.98% |
| 2018-11 | 1187823 | 1331601 | 12.10% | 1271709 | | 7.06% |
| 2018-12 | 1228834 | 1364951 | 11.08% | 1259992 | | 2.54% |
| **Average 2018** | | **11.51%** |  |  | | **4.13%** |

Again, all models fit the data very well, as shown by the high R2. The values it takes are greater than 0.97 in all cases.

The results demonstrate the superiority of Exponential Smoothing over SARMA. With the exception of the first test, HW outperforms SARMA (3,4) (1,1) in all other cases. In test 1, MAPE is slightly lower for SARMA (7.92% versus 9.02%), but as the samples get larger the HW approach becomes more accurate. For the 2014 and 2015 forecasts the SARMA’s MAPE is almost twice as high as that of HW.

Importantly, Exponential Smoothing manages to capture the upward trend in nights-total. Focusing on August values, they rose by 2% year-over-year from 2013 to 2014 and 5.1% from 2014 to 2015. While SARMA correctly predicts the rising trend from August 2013 to August 2014, it fails to identify that this trend would continue in August 2015 and mistakenly predicts a lower value. On the contrary, the HW approach is successful in tracking those changes.

The above results provide evidence that Exponential Smoothing can yield more reliable predictions for this specific variable than SARMA. Therefore, the out-of-sample forecasts of HW are perceived as more realistic in comparison to the respective figures of SARMA. Indeed, the 7.64% increase that HW predicts for 2016 and the 11.51% for 2017 are much closer to industry expectations than the 3.52% and 4.13% of SARMA.

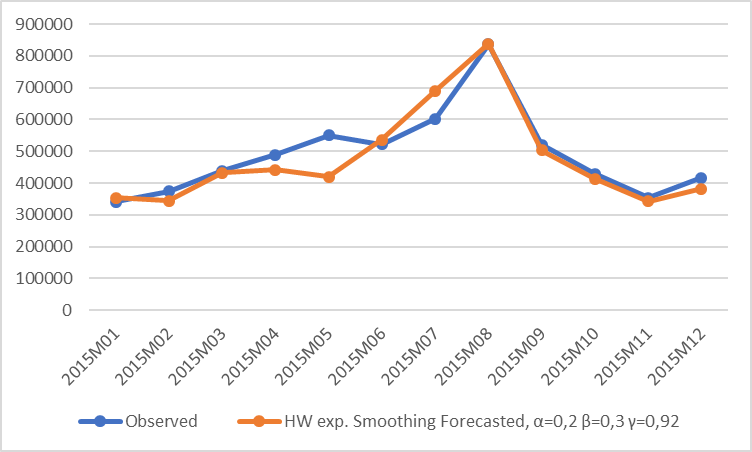
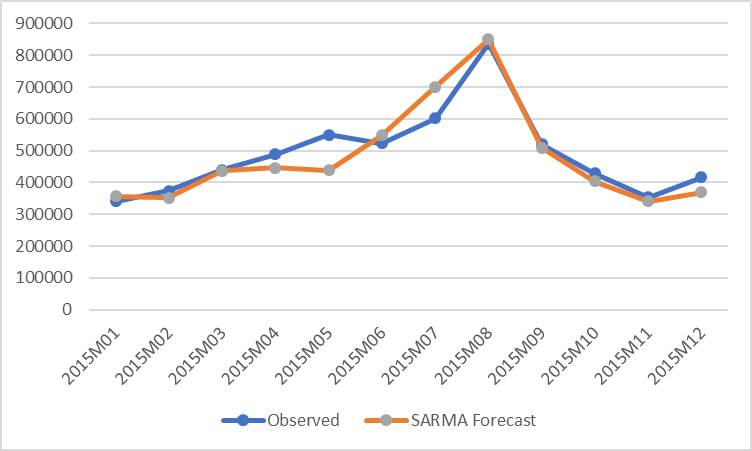
### **3.6.3. Arrivals – Residents – Total**

Arrivals – Residents are divided into the sub-sample 2000 – 2014, with the forecast period being 2015. In addition, we produce out-of-sample forecast for 2016 – 2018.

Below are the forecasting tests.

**TEST 1: 2000 - 2014 => 2015**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Period No | Month | Observed | HW Forecast | SARMA Forecast |
| 181 | 2015M01 | 340579 | 353927 | 357094 |
| 182 | 2015M02 | 374333 | 344732 | 351473 |
| 183 | 2015M03 | 439043 | 432265 | 436433 |
| 184 | 2015M04 | 488354 | 442521 | 446098 |
| 185 | 2015M05 | 550000 | 420032 | 437993 |
| 186 | 2015M06 | 522652 | 536087 | 548808 |
| 187 | 2015M07 | 601938 | 690098 | 700226 |
| 188 | 2015M08 | 836737 | 838141 | 850680 |
| 189 | 2015M09 | 520410 | 502712 | 508786 |
| 190 | 2015M10 | 429080 | 413758 | 403938 |
| 191 | 2015M11 | 353463 | 342972 | 341023 |
| 192 | 2015M12 | 415984 | 382023 | 369585 |
| **Errors :** |  |  |  |  |
| **MAPE** |  |  | **6.82%** | **7.19%** |
| **MAD** |  |  | **33833** | **35853** |
| **MADP** |  |  | **6.9%** | **0.00021%** |
| **MSD** |  |  | **2488238280** | **2395613015** |
| **R2** |  |  | **0.8540** | **0.8594** |

**TEST 2: 2000 - 2015 => 2016, 2017, 2018**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **HW** | | **SARMA** | |
| **Month** | Observed 2015 | Forecast | % Difference | Forecast | % Difference |
| 2016-01 | 340579 | 364921 | 7.15% | 386265 | 13.41% |
| 2016-02 | 374333 | 391100 | 4.48% | 386495 | 3.25% |
| 2016-03 | 439043 | 459301 | 4.61% | 450467 | 2.60% |
| 2016-04 | 488354 | 498027 | 1.98% | 501970 | 2.79% |
| 2016-05 | 550000 | 524798 | -4.58% | 517878 | -5.84% |
| 2016-06 | 522652 | 512161 | -2.01% | 546431 | 4.55% |
| 2016-07 | 601938 | 611719 | 1.62% | 684422 | 13.70% |
| 2016-08 | 836737 | 847980 | 1.34% | 867190 | 3.64% |
| 2016-09 | 520410 | 527809 | 1.42% | 525280 | 0.94% |
| 2016-10 | 429080 | 436502 | 1.73% | 442503 | 3.13% |
| 2016-11 | 353463 | 361152 | 2.18% | 358914 | 1.54% |
| 2016-12 | 415984 | 420265 | 1.03% | 402733 | -3.19% |
| **Average 2016** | |  | **1.75%** |  | **3.38%** |
| 2017-01 | 340579 | 369883 | 8.60% | 388191 | 13.98% |
| 2017-02 | 374333 | 396412 | 5.90% | 385937 | 3.10% |
| 2017-03 | 439043 | 465532 | 6.03% | 452723 | 3.12% |
| 2017-04 | 488354 | 504776 | 3.36% | 508792 | 4.19% |
| 2017-05 | 550000 | 531902 | -3.29% | 514934 | -6.38% |
| 2017-06 | 522652 | 519086 | -0.68% | 548528 | 4.95% |
| 2017-07 | 601938 | 619981 | 3.00% | 689742 | 14.59% |
| 2017-08 | 836737 | 859420 | 2.71% | 856762 | 2.39% |
| 2017-09 | 520410 | 534921 | 2.79% | 527985 | 1.46% |
| 2017-10 | 429080 | 442378 | 3.10% | 447212 | 4.23% |
| 2017-11 | 353463 | 366008 | 3.55% | 358126 | 1.32% |
| 2017-12 | 415984 | 425909 | 2.39% | 406306 | -2.33% |
| **Average 2017** | |  | **3.12%** |  | **3.72%** |
| 2018-01 | 340579 | 374845 | 10.06% | 392119 | 15.13% |
| 2018-02 | 374333 | 401725 | 7.32% | 384817 | 2.80% |
| 2018-03 | 439043 | 471764 | 7.45% | 456339 | 3.94% |
| 2018-04 | 488354 | 511526 | 4.74% | 511400 | 4.72% |
| 2018-05 | 550000 | 539006 | -2.00% | 511929 | -6.92% |
| 2018-06 | 522652 | 526011 | 0.64% | 551843 | 5.59% |
| 2018-07 | 601938 | 628243 | 4.37% | 689703 | 14.58% |
| 2018-08 | 836737 | 870860 | 4.08% | 847267 | 1.26% |
| 2018-09 | 520410 | 542034 | 4.16% | 531445 | 2.12% |
| 2018-10 | 429080 | 448253 | 4.47% | 448727 | 4.58% |
| 2018-11 | 353463 | 370864 | 4.92% | 357932 | 1.26% |
| 2018-12 | 415984 | 431553 | 3.74% | 410151 | -1.40% |
| **Average 2018** | |  | **4.50%** |  | **3.97%** |

The R-squared is satisfactory for both models, as it is greater than 0.85. Thus, the two models are acceptable in terms of goodness of fit.

The forecasting results show that Exponential Smoothing performs slightly better than SARMA (3,4) (1,1) in the context of arrivals of non-residents. MAPE is lower by 0.37% for HW over SARMA. MAD is also lower by 2,020 points. On the other hand, MADP and MSD suggest that SARMA is more accurate than Exponential Smoothing. Overall, the two models have an almost equivalent performance. Yet, the greater reliability of MAPE and MAD measures, coupled with the better performance of HW in the preceding forecasts confer an edge on Exponential Smoothing over SARMA. From this perspective, the out-of-sample forecasts generated by HW are deemed more credible. However, the predictions of the two approaches for 2017 and 2018 do not differ significantly. Based on their forecasts, the expected average growth rate for 2017 is 3.1 – 3.7% and for 2018, 4.0 – 4.5%.

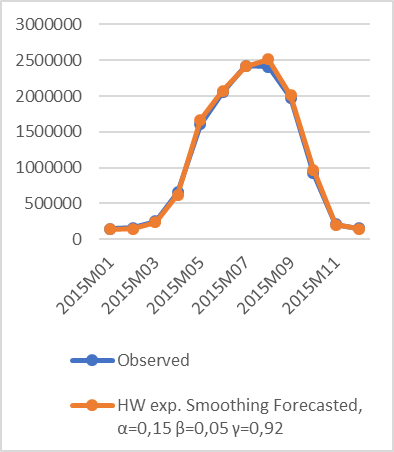
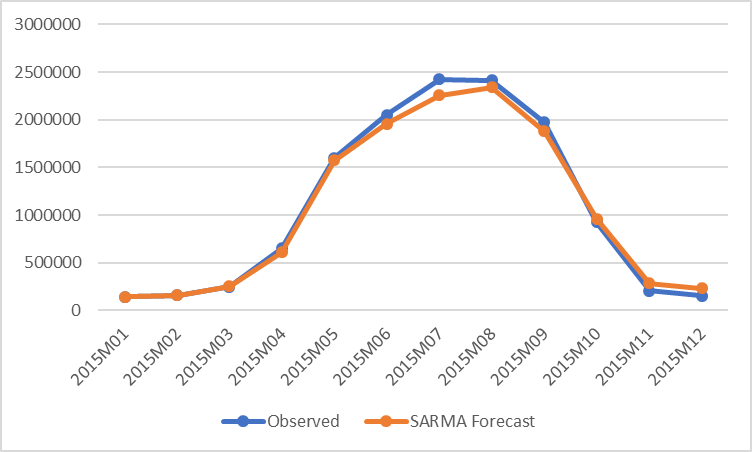
### **3.6.4. Arrivals – Non-Residents – Total**

This time-series is divided into the sub-sample 2000 – 2014, with the forecast period being 2015. In addition, we conduct out-of-sample analysis for 2016 – 2018.

Below are the forecasting tests.

**TEST 1: 2000 - 2014 => 2015**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Period No | Month | Observed | HW Forecast | SARMA Forecast |
| 181 | 2015M01 | 144852 | 137780 | 144291 |
| 182 | 2015M02 | 158260 | 147852 | 161799 |
| 183 | 2015M03 | 250462 | 237190 | 254443 |
| 184 | 2015M04 | 653432 | 623234 | 609026 |
| 185 | 2015M05 | 1602897 | 1669301 | 1575870 |
| 186 | 2015M06 | 2051321 | 2075122 | 1959918 |
| 187 | 2015M07 | 2423206 | 2414304 | 2256429 |
| 188 | 2015M08 | 2410735 | 2514710 | 2338328 |
| 189 | 2015M09 | 1972763 | 2009524 | 1881238 |
| 190 | 2015M10 | 922342 | 972835 | 962135 |
| 191 | 2015M11 | 205587 | 204405 | 283498 |
| 192 | 2015M12 | 152749 | 148314 | 233731 |
| **Errors :** |  |  |  |  |
| **MAPE** |  |  | **3.51%** | **10.58%** |
| **MAD** |  |  | **29742** | **58359** |
| **MADP** |  |  | **2.8%** | **0.0000071%** |
| **MSD** |  |  | **1752865401** | **5560950618** |
| **R2** |  |  | **0.9979** | **0.9932** |

**TEST 2: 2000 - 2015 => 2016, 2017, 2018**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **HW** | | **SARMA** | |
| **Month** | Observed 2015 | Forecast | % Difference | Forecast | % Difference |
| 2016-01 | 144852 | 153891 | 6.24% | 166707 | 15.09% |
| 2016-02 | 158260 | 166324 | 5.10% | 199781 | 26.24% |
| 2016-03 | 250462 | 261816 | 4.53% | 303747 | 21.27% |
| 2016-04 | 653432 | 680248 | 4.10% | 717554 | 9.81% |
| 2016-05 | 1602897 | 1693689 | 5.66% | 1683591 | 5.03% |
| 2016-06 | 2051321 | 2168853 | 5.73% | 2123158 | 3.50% |
| 2016-07 | 2423206 | 2560457 | 5.66% | 2481637 | 2.41% |
| 2016-08 | 2410735 | 2573971 | 6.77% | 2494220 | 3.46% |
| 2016-09 | 1972763 | 2107967 | 6.85% | 2047618 | 3.79% |
| 2016-10 | 922342 | 995354 | 7.92% | 1029734 | 11.64% |
| 2016-11 | 205587 | 220401 | 7.21% | 313747 | 52.61% |
| 2016-12 | 152749 | 162633 | 6.47% | 262594 | 71.91% |
| **Average 2016** | |  | **6.02%** |  | **18.90%** |
| 2017-01 | 144852 | 164172 | 13.34% | 271976 | 87.76% |
| 2017-02 | 158260 | 177375 | 12.08% | 301762 | 90.68% |
| 2017-03 | 250462 | 279115 | 11.44% | 400918 | 60.07% |
| 2017-04 | 653432 | 724946 | 10.94% | 813502 | 24.50% |
| 2017-05 | 1602897 | 1804374 | 12.57% | 1775001 | 10.74% |
| 2017-06 | 2051321 | 2309824 | 12.60% | 2217457 | 8.10% |
| 2017-07 | 2423206 | 2725984 | 12.49% | 2577429 | 6.36% |
| 2017-08 | 2410735 | 2739480 | 13.64% | 2587601 | 7.34% |
| 2017-09 | 1972763 | 2242789 | 13.69% | 2142218 | 8.59% |
| 2017-10 | 922342 | 1058677 | 14.78% | 1120761 | 21.51% |
| 2017-11 | 205587 | 234349 | 13.99% | 404607 | 96.81% |
| 2017-12 | 152749 | 172871 | 13.17% | 355011 | 132.41% |
| **Average 2017** | |  | **12.89%** |  | **46.24%** |
| 2018-01 | 144852 | 174453 | 20.44% | 365174 | 152.10% |
| 2018-02 | 158260 | 188425 | 19.06% | 397174 | 150.96% |
| 2018-03 | 250462 | 296413 | 18.35% | 497147 | 98.49% |
| 2018-04 | 653432 | 769645 | 17.79% | 911419 | 39.48% |
| 2018-05 | 1602897 | 1915060 | 19.47% | 1873704 | 16.89% |
| 2018-06 | 2051321 | 2450794 | 19.47% | 2317022 | 12.95% |
| 2018-07 | 2423206 | 2891512 | 19.33% | 2676799 | 10.47% |
| 2018-08 | 2410735 | 2904989 | 20.50% | 2688650 | 11.53% |
| 2018-09 | 1972763 | 2377611 | 20.52% | 2243184 | 13.71% |
| 2018-10 | 922342 | 1122001 | 21.65% | 1223545 | 32.66% |
| 2018-11 | 205587 | 248296 | 20.77% | 507611 | 146.91% |
| 2018-12 | 152749 | 183109 | 19.88% | 458916 | 200.44% |
| **Average 2018** | |  | **19.77%** |  | **73.88%** |

The R-squared is higher than 0.99 for both models. Therefore, no concerns are raised about the goodness of fit.

In the 2015 forecasting exercise, SARIMA (3,1,3) (1,1,1) performs rather poorly compared to Exponential Smoothing. Its MAPE is about three time larger than HW’s (10.6% versus 3.5%). This is consistent with the findings for the rest of the variables examined in this study.

The proposed SARIMA model has been adjusted in such a way as to capture the trend exhibited by this variable. Yet, it seems that the model overpredicts the tourist arrivals by residents and this causes this large error. This tendency passes on to out-of-sample forecasts and yield very high growth rates (i.e. 46.2% for 2017 and 73.9% for 2018). On the other hand, HW produces more conservative forecasts for those years (12.9% for 2017 and 19.8% for 2018).

## **3.7. Conclusion**

The preceding analysis suggests that both Exponential Smoothing and SAR(I)MA are powerful modelling approaches that can generate satisfactory predictions for the future tourism demand in Greece. The R-squared values are greater than 0.95 in most cases, indicating that the proposed models fit very well the sets of observations.

The out-of-sample forecasts for the period 2016 – 2018 reveal a growing rate for all variables of this analysis. In fact, there is consensus between the two modelling approaches that a booming tourism demand is anticipated during this period. This is in line with industry expectations and press announcements that point to higher tourism flows to Greece.

The comparison of pertinent forecast errors shows that Triple Exponential Smoothing manages to predict the future values of almost all variables with remarkable accuracy. On several occasions, MAPE is as low as 2.5 - 3.5%. At the other edge of the spectrum, while in certain cases SAR(I)MA reaches a satisfactory MAPE of about 3.5%, in other cases MAPE exceeds 8%.

Overall, the multiple comparisons illustrated that HW outperforms SAR(I)MA in the majority of forecasting scenarios. These findings contradict several studies from the tourism literature, according to which the ARIMA class of models performs better than exponential smoothing. The outcome of the various forecasting exercises conducted within this study is that the HW approach is more suitable for the Greek tourism data. This is in line with the fact that in cases where the structure of the data generating process is changing – as in the dataset of this study -, HW is more appropriate than SAR(I)MA.

Therefore, Triple Exponential Smoothing is proposed as the major forecasting approach of variables pertaining to Greek tourism demand. On the other hand, SAR(I)MA, although less accurate, is still a robust modelling technique and can be used as a complement in order to enhance the reliability of forecasts. A further alternative could be the combined forecasting based on both the Triple Exponential Smoothing and the SAR(I)MA technique.

# **Final remarks**

To finalize the report, a summary is given about the analyzed data as well about the analysis of forecasting methods. The paper is closed with conclusions which also give an outlook on the possible operation of forecasting the Greek tourism in the future, based on the results of this project.

## **4.1. Summary**

**The data set**: This study is based on a data set of monthly observations, provided by ELSTAT. The focal point of the analysis includes the number of arrivals (in hotels & campsites, total, residents and non-residents) and the number of overnights (in hotels & campsites, total, residents and non-residents).

The analysis of the behavior of each variable indicates that in most cases there is a slight positive trend without any significant changes. The shape of variation over time diverges from linear trend. The number of overnights, total, (variable Nights-Total) behaves differently, as it exhibits a negative trend for three months. At the same time, the projection of the linear trend line in the future period 2017-2018 gives an indication of what will follow in the coming months. This of course is investigated more thoroughly in the forecasting approaches of section 3.

The next step involved the analysis by seasons, which justified the categorization into three groups of months: November - March (Low Season), October and April (Intermediate Season), and May - September (High Season). The mean and Anova tests show that the mean values of the variables Nights-Total and Arrivals-Total in the three seasons are significantly different, confirming the distinction between them. The ensuing correlation analysis showed that there is a high linear correlation between Nights-Total and Arrivals-Total (R2 = 0,98). It also arises that such strong correlation is not present when it comes to composition factors, such as Campsites, Residents and Non-Residents.

The most significant origin countries in terms of number of arrivals are found to be Germany with a share of 15,01%, United Kingdom with 13,09%, France with 6,99%, and Italy with 6,54%. The majority of EU countries, such as Germany, UK, France, Italy, Sweden, Switzerland, Netherlands, Belgium, and additionally Cyprus, Serbia-Montenegro, and USA appear to have a relatively constant flow of arrivals. Balkan countries (Bulgaria, Albania, Turkey), Poland, and Russia exhibit a much higher variation of arrivals.

The analysis of trend lines by country of origin indicated that Italy, UK, Germany, Netherlands, USA, Belgium, and Sweden have an almost zero trend, Russia, Poland, Albania, France, Turkey, Switzerland, Serbia-Montenegro, Romania, and Bulgaria a positive trend, and finally Cyprus a negative trend.

It is noteworthy that time-series patterns vary significantly from country to country. The categorization in seasons made for the total number of arrivals does not match the pattern observed in the arrivals from certain countries, such as Albania, Bulgaria, and Turkey. On top of that, while the peak in total arrivals occurs in August, this is not always the case with arrivals from other countries, such as Germany and the US.

**The forecasting analysis** demonstrated that both Exponential Smoothing and SAR(I)MA are robust modeling approaches that can yield satisfactory predictions for the variables examined. The high R-squared values (above 0.95 in most cases) provide evidence that the data fit the model very well.

The evaluation of the forecasting accuracy of each model revealed that Triple Exponential Smoothing outperforms SAR(I)MA in most of the tests. This shows that the type of data used in this study can be modeled more effectively by means of Exponential Smoothing. Yet, ARIMA class of models remains a useful forecasting instrument and this is made evident not only by the findings of the relevant literature, but also by the forecast errors of the present study. Those errors may be higher than the ones generated by HW, but they remain within acceptable limits. In a nutshell, the HW approach is proven superior for this kind of analysis, but that is not to say that SARIMA should be discarded. In fact, a combination of the two methods may provide a more comprehensive picture of what lies ahead.

## **4.2. Conclusions and next steps**

The intention of the forecasting project was to analyze forecasting methods with the aim (1) to find the suitable methodology for the routine-wise implementation of forecasting key statistics of Greek tourism in the future and (2) to make a proposal on how to develop a tool for forecasting the tourism development both as a policy tool for the Ministry of Tourism and as a campaigning tool for the Greek National Tourism Organization (GNTO). The project has been conducted in cooperation with two Greek experts, Dr. Yiannis Smirlis (University of Piraeus) and Dr. Vangelis Tsioumas (The American College of Greece - DEREE) who will be available for the further development of the proposed methodology into a routine-wise operated forecasting tool.

This part of the report discusses the next steps to be taken in order to transform the proposed forecasting methodology in a routine-wise used tool: This tool shall annually or sub-annually provide the Ministry of Tourism and the GNTO, but also the businesses, the media, and the public of Greece with forecasts which are needed for policy planning and developing campaigns in Greek tourism.

**The institutional framework**: For the future routine-wise operation of the proposed forecasting methods, an institutional arrangement has to be established. The institution which shall have the responsibility for the future operation has to deal with the following tasks:

* Development of an IT system which allows handling the data and estimating the forecasts
* Establishing the data base for the forecasting system, and updating and editing the data
* Running the forecasting procedures according to an agreed schedule
* Production and dissemination of the forecasting reports

**The data base**: The forecasting project was restricted in the scope of analysis by the availability of tourism data. The main source of tourism data is ELSTAT; the available data were the number of arrivals and the number of overnights in hotels etc. and campsites, with breakdowns corresponding to inbound and domestic visitors and others. These data allow forecasts of certain tourism flows. Further areas of interest for the project are expenditures for tourism consumption, as well as economic determinants of inbound flows from certain countries of residence such as the effect of changes in the disposable income of the visitors, the price levels in Greece and alternative destinations, the travel costs, but also factors of the economic environment like the exchange rates, and factors like the Greek marketing expenditures, events like political unrests, and others.

Based on the experience of the forecasting project, some recommendations are provided that may help to improve the data base for the future forecasting system:

* The project was based on the most recent data available: The time-series were – in late spring of 2017 – available up to the end of 2015. For forecasting tourism flows in 2017 in reasonable quality, input data up to the end of 2016 should be available.
* ELSTAT’s data should be provided in a form that will not require time-consuming transformations each time they are to be used. Downloading the time-series from the ELSTAT-website should be possible. A suitable format is proposed in section 2.1 of the report.
* The available data need to be enriched with the number of overnights by country of origin, at least for the important source countries of Greek tourism. This will allow a more complete analysis of the tourism flows and to this extent the development of region-specific forecasting models.
* ELSTAT’s data base should be supplemented (1) by regional data (NUTS 2 level), at least for the tourism-relevant regions, as well as (2) by the expenditures of tourists at a reasonable level of detail.
* The consistency of the data provided by ELSTAT with corresponding data from other data sources like Eurostat and SETE and of the definitions given in the metadata reports of all these agencies should be checked; discrepancies should be documented.

**The forecasting methods:** The project report generally surveys forecasting methods which are reported in the statistical literature as relevant for tourism forecasting. Given the fact that only tourism flow data were available for the empirical part of the forecasting project, the range of methods which could be analysed was rather limited. Due to the lack of data and also of time, the involvement was also restricted with respect to the used statistical method.

The main result of the forecasting project is the demonstration that for tourism flows (number of arrivals, numbers of overnights), exponential smoothing and also the SAR(I)MA technique can yield satisfactory predictions. Given the fact that tourism flows and their forecasts are an important and crucial ingredient for policy planning and for developing campaigns, the results of the project are a reasonable starting point for the future routine-wise operation of the forecasting methods.

Methods that are based on more sophisticated models like Vector Autoregressive models, Error Correction models, Time Varying Parameter (TVP) models, etc. allow for deeper analyses of Greek tourism, in particular the analysis of the economic background of the tourism key statistics. For example, analysing the economic determinants of inbound flows from certain countries of residence such as the effect of changes in the disposable income of the visitors, the price levels in Greece and alternative destinations, the travel costs, but also factors of the economic environment like the exchange rates, and factors like the Greek marketing expenditures, events like political unrests, and others may result in more sophisticated models which allow for assessing the effects of changes in these economic determinants on the inbound flows; the forecasts may, e.g., reflect the increased travel costs or the reduced unemployment rate in Germany on the inflow of German visitors. The further development of forecasting models should be in the responsibility of the institution which shall deal with the future routine-wise operation of forecasting Greek tourism; cooperation with research institutions is highly recommended for this purpose.

# **Appendices**

## **Appendix I. Descriptive Statistics of all variables included in the data set**



Figure II-4 Descriptive Statistics for Arrivals-Total



Figure II-5 Descriptive Statistics for Arrivals-Campsites-Total



Figure II-6 Descriptive Statistics for Arrivals-Hotels-Total



Figure II-7 Descriptive Statistics for Arrivals-Non-Residents-Total



Figure II-8 Descriptive Statistics for Arrivals-Campsites-Non-Residents



Figure II-9 Descriptive Statistics for Arrivals-Hotels-Non-Residents



Figure II-10 Descriptive Statistics for Arrivals-Residents-Total



Figure II-11 Descriptive Statistics for Arrivals-Campsites-Residents



Figure II-12 Descriptive Statistics for Arrivals-Hotels-Residents



Figure II-13 Descriptive Statistics for Nights-Total



Figure II-14 Descriptive Statistics for Nights-Campsites-Total



Figure II-15 Descriptive Statistics for Nights-Hotels-Total



Figure II-16 Descriptive Statistics for Nights-Non-Residents-Total



Figure II-17 Descriptive Statistics for Nights-Campsites-Non-Residents



Figure II-18 Descriptive Statistics for Nights-Hotels-Non-Residents



Figure II-19 Descriptive Statistics for Nights-Residents-Total



Figure II-20 Descriptive Statistics for Nights-Campsites-Residents



Figure II-21 Descriptive Statistics for Nights-Hotels-Residents

## **Appendix II. Plots of Nights-Total, Arrivals-Total by Month**

###### **01 January, Nights**

###### 

###### Figure II-1 January, Linear Trend for Nights-Total

###### 

###### Figure II-2 January, Statistical Summary for Nights-Total

###### 

###### Figure II‑3 January, break down time-series plot for Nights-Total

###### **01 January, Arrivals**

###### 

###### Figure II-4 January, Linear Trend for Arrivals-Total

###### 

###### Figure II-6 January, Statistical Summary for Arrivals-Total

###### 

###### Figure II-7 January, break down time-series plot for Arrivals-Total

###### **02 February, Nights**

###### 

###### Figure II-8 February, Linear Trend for Nights-Total

###### 

###### Figure II-9 February, Statistical Summary for Nights-Total

###### 

###### Figure II-10 February, break down time-series plot for Nights-Total

###### **02 February, Arrivals**

###### 

###### Figure II-11 February, Linear Trend for Arrivals-Total

###### 

###### Figure II-12 February, Statistical Summary for Arrivals-Total

###### 

###### Figure II-13 February, break down time-series plot for Arrivals-Total

###### **03 March, Nights**

###### 

###### Figure II-14 March, Linear Trend for Nights-Total

###### 

###### Figure II-15 March, Statistical Summary for Nights-Total

###### 

###### Figure II-16 March, break down time-series plot for Nights-Total

###### **03 March Arrivals**

###### 

###### Figure II-17 March, Linear Trend for Arrivals-Total

###### 

###### Figure II-18 March, Statistical Summary for Arrivals-Total

###### 

###### Figure II-19 March, break down time-series plot for Arrivals-Total

###### **04 April, Nights**

###### 

###### Figure II-20 April, Linear Trend for Nights-Total

###### 

###### Figure II-21 April, Statistical Summary for Nights-Total

###### 

###### *Figure II-22 April, break down time-series plot for Nights-Total*

###### **04 April, Arrivals**

###### 

###### Figure II-23 *April*, Linear Trend for Arrivals-Total

###### 

###### Figure II-24 April, Statistical Summary for Arrivals-Total

###### 

###### Figure II-25 April, break down time-series plot for Arrivals-Total

###### **05 May, Nights**

###### 

###### Figure II-26 May, Linear Trend for Nights-Total

###### 

###### Figure II-27 May, Statistical Summary for Nights-Total

###### 

###### Figure II-28 May, break down time-series plot for Nights-Total

###### **05 May, Arrivals**

###### 

###### Figure II-29 May, Linear Trend for Arrivals-Total

###### 

###### Figure II-30 May, Statistical Summary for Arrivals-Total

###### 

###### Figure II-31 May, break down time-series plot for Arrivals-Total

###### **06 June, Nights**

###### 

###### Figure II-32 June, Linear Trend for Nights-Total

###### 

###### Figure II-33 June, Statistical Summary for Nights-Total

###### 

###### Figure II-34 June, break down time-series plot for Nights-Total

###### **06 June, Arrivals**

###### 

###### Figure II-35 June, Linear Trend for Arrivals-Total

###### 

###### Figure II-36 June, Statistical Summary for Arrivals-Total

###### 

###### Figure II-37 June, break down time-series plot for Arrivals-Total

###### **07 July, Nights**

###### 

###### Figure II-38 July, Linear Trend for Nights-Total

###### 

###### Figure II-39 July, Statistical Summary for Nights-Total

###### 

###### Figure II-40 July, break down time-series plot for Nights-Total

###### **07 July, Arrivals**

###### 

###### Figure II-41 July, Linear Trend for Arrivals-Total

###### 

###### Figure II-42 July, Statistical Summary for Arrivals-Total

###### 

###### Figure II-43 July, break down time-series plot for Arrivals-Total

###### **08 August, Nights**

###### 

###### Figure II-44 August, Linear Trend for Nights-Total

###### 

###### Figure II-45 August, Statistical Summary for Nights-Total

###### 

###### Figure II-46 August, break down time-series plot for Nights-Total

###### **08 August, Arrivals**

###### 

###### Figure II-47 August, Linear Trend for Arrivals-Total

###### 

###### Figure II-48 August, Statistical Summary for Arrivals-Total

###### 

###### Figure II-49 August, break down time-series plot for Arrivals-Total

###### **09 September, Nights**

###### 

###### Figure II-50 September, Linear Trend for Nights-Total

###### 

###### Figure II-51 September, Statistical Summary for Nights-Total

###### 

###### Figure II-52 September, break down time-series plot for Nights-Total

###### **09 September, Arrivals**

###### 

###### Figure II-53 September, Linear Trend for Arrivals-Total

###### 

###### Figure II-54 September, Statistical Summary for Nights-Total

###### 

###### Figure II-55 September, break down time-series plot for Nights-Total

###### **10 October, Nights**

###### 

###### Figure II-56 October, Linear Trend for Nights-Total

###### 

###### Figure II-57 October, Statistical Summary for Nights-Total

###### 

###### Figure II-58 October, break down time-series plot for Nights-Total

###### **10 October, Nights**

###### 

###### Figure II-59 October, Linear Trend for Arrivals-Total

###### 

###### Figure II-60 October, Statistical Summary for Arrivals-Total

###### 

###### Figure II-61 October, break down time-series plot for Nights-Total

###### **11 November, Nights**

###### 

###### Figure II-62 November, Linear Trend for Nights-Total

###### 

###### Figure II-63 November, Statistical Summary for Nights-Total

###### 

###### Figure II-64 November, break down time-series plot for Nights-Total

###### **11 November, Arrivals**

###### 

###### Figure II-65 November, Linear Trend for Arrivals-Total

###### 

###### Figure II-66 November, Statistical Summary for Arrivals-Total

###### 

###### Figure II-67 November, break down time-series plot for Arrivals-Total

###### **12 December, Nights**

###### 

###### Figure II-68 December, Linear Trend for Nights-Total

###### 

###### Figure II-69 December, Statistical Summary for Nights-Total

###### 

###### Figure II-70 December, break down time-series plot for Nights-Total

###### **12 December, Arrivals**

###### 

###### Figure II-71 December, Linear Trend for Arrivals-Total

###### 

###### Figure II-72 December, Statistical Summary for Arrivals-Total

###### 

###### Figure II-73 December, break down time-series plot for Arrivals-Total

## **Appendix III. Mean and ANOVA tests for Seasons**

**Mean and Anova tests**

Mean Tests

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **One-Sample Test** | | | | | | |
|  | Test Value = 0 | | | | | |
| t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Arrivals-Campsites-Total | 8.970 | 191 | .000 | 27812.219 | 21696.24 | 33928.20 |
| Arrivals-Hotels-Total | 23.114 | 191 | .000 | 1223552.307 | 1119139.38 | 1327965.24 |
| Nights-Hotels-Total | 16.116 | 191 | .000 | 5185974.151 | 4551265.62 | 5820682.68 |
| Nights-Campsites-Total | 7.667 | 191 | .000 | 95511.755 | 70938.68 | 120084.83 |

**One-way ANOVA: Nights-Total versus Season**

Source DF SS MS F P

Season 2 3,16183E+15 1,58092E+15 352,73 0,000

Error 189 8,47097E+14 4,48199E+12

Total 191 4,00893E+15

S = 2117072 R-Sq = 78,87% R-Sq(adj) = 78,65%

Individual 95% CIs For Mean Based on

Pooled StDev

Level N Mean StDev --------+---------+---------+---------+-

1 80 9638758 3172561 (-\*)

2 32 3235670 1128468 (--\*--)

3 80 904806 397362 (-\*)

--------+---------+---------+---------+-

2500000 5000000 7500000 10000000

Pooled StDev = 2117072

Grouping Information Using Tukey Method

Season N Mean Grouping

1 80 9638758 A

2 32 3235670 B

3 80 904806 C

Means that do not share a letter are significantly different.

Tukey 95% Simultaneous Confidence Intervals

All Pairwise Comparisons among Levels of Season

Individual confidence level = 98,08%

Season = 1 subtracted from:

Season Lower Center Upper -------+---------+---------+---------+--

2 -7448906 -6403088 -5357270 (--\*--)

3 -9524516 -8733952 -7943388 (-\*-)

-------+---------+---------+---------+--

-7000000 -3500000 0 3500000

Season = 2 subtracted from:

Season Lower Center Upper -------+---------+---------+---------+--

3 -3376683 -2330865 -1285047 (--\*--)

-------+---------+---------+---------+--

-7000000 -3500000 0 3500000

**One-way ANOVA: Arrivals-Total versus Season**

Source DF SS MS F P

Season 2 9,34028E+13 4,67014E+13 448,90 0,000

Error 189 1,96625E+13 1,04034E+11

Total 191 1,13065E+14

S = 322543 R-Sq = 82,61% R-Sq(adj) = 82,43%

Individual 95% CIs For Mean Based on Pooled StDev

Level N Mean StDev -+---------+---------+---------+--------

1 80 2050593 485580 (\*)

2 32 1040318 118331 (-\*-)

3 80 536555 87233 (-\*)

-+---------+---------+---------+--------

500000 1000000 1500000 2000000

Pooled StDev = 322543

Grouping Information Using Tukey Method

Season N Mean Grouping

1 80 2050593 A

2 32 1040318 B

3 80 536555 C

Means that do not share a letter are significantly different.

Tukey 95% Simultaneous Confidence Intervals

All Pairwise Comparisons among Levels of Season

Individual confidence level = 98,08%

Season = 1 subtracted from:

Season Lower Center Upper -------+---------+---------+---------+--

2 -1169609 -1010275 -850941 (-\*--)

3 -1634483 -1514038 -1393593 (-\*-)

-------+---------+---------+---------+--

-1200000 -600000 0 600000

Season = 2 subtracted from:

Season Lower Center Upper -------+---------+---------+---------+--

3 -663097 -503763 -344429 (--\*-)

-------+---------+---------+---------+--

-1200000 -600000 0 600000

## **Appendix IV. Literature Review – Tourism Forecasting**

*Table 1. Publications based on ARIMA models*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reference | Model type | Dependent variable(s) | Independent variable(s) | Data | Comparison with | Outcome |
| Kim & Uysal(1998) | ARMAX | Hotels in Seoul,Korea -number of nights | Price of hotel rooms, number of events (e.g. international conventions, seminars and exhibitions), and total trade volume of the destination country | Monthly - 1/1991 -12/1995 | - | Price, number of events and trade volume directly affected demand |
| Cho (2001) | ARIMAX (adjusted ARIMA) | Hong Kong tourist arrivals | ARIMA residuals of GDP/GNP, unemployment rate, money supply, Consumer Price Index (CPI), imports, exports, and discount rate | Quarterly -  1/1975 -4/1997 | (univariate) ARIMA, Exponential Smoothing | ARIMAX is the best forecasting method for Japan, while univariate ARIMA is the best predictor for the US and UK |
| Akal (2004) | AR(I)MAX | Tourism revenues for Turkey | Tourist arrivals | Annual - 1963 - 2001 | Static Regression | AR(I)MAX outperformed the simple econometric cause–effect technique |
| Gounopoulos, Petmezas, & Santamaria (2012) (\*) | ARIMA (1, 1, 1) | Tourist numbers to Greece | Unemployment levels, Consumer Price Indices (CPI), and Consumer Confidence Indicators of the country of origin | Monthly - 1/1977-12/2009 | Exponential smoothing models | ARIMA(1,1,1) performs better |
| ShuiKi, ShinHuei, & ChiKeung (2013) | SARIMA | Tourist arrivals to Hong Kong |  | Monthly | Seasonal moving average model &  Holt-Winter model. | SARIMA pefrorms better |
| Baldigara and Mamula (2015) | SARIMA | German tourist flows to Croatia |  | Quartlerly - 2003 - 2012 | No comparison | ARIMA(0,0,0)(1,1,3) is adequate |
| Hassani, Silva, Antonakakis, Filis, & Gupta (2017) | Univariate approach / Parametric and nonparametric forecasting techniques, including ARIMA | international tourist arrivals to European countries |  |  | ETS, NN, TBATS, ARFIMA, MA, WMA, SSA-R and SSA-V | No single model outperforms others in all cases. Neural Networks and ARFIMA have the worst performance |

(\*) Gounopoulos, Petmezas, & Santamaria (2012) use of unemployment and consumer confidence as alternative proxies for the state of the economy in the country of origin. Although no studies have yet to document the impact of macroeconomic shocks on future tourists’ arrivals, reviewing the wider literature on the impact of unemployment, changes in the tourists’ cost of living, and consumer confidence could provide useful inferences on potential effects to international tourism flows.

**Causal forecasting**

A number of authors put forward causal forecasting approaches with the aim of predicting tourism demand. Their studies are based on the selection of a set of factors they consider as most influential. Those factors are fed into pertinent models and generate forecasts. As discussed in the ensuing literature review, the most widely used models are the ADLM, the VAR and the ECM. The performance of each technique is assessed by means of comparison with alternative and benchmark models.

Witt & Witt (1992) employ the ADLM approach for the case of tourist arrivals. The explanatory variables under consideration include income in origin country, prices, exchange rate, travel cost (airfare), travel cost to substitute destinations, currency restrictions, and dummy variables. In addition they apply univariate forecasting methods, such as AR, ARIMA, and Exponential Smoothing. Syriopoulos (1995) adopt an ECM approach to estimate tourism expenditure. The explanatory variable of this model include income in origin country, relative tourism price, exchange rate, and dummy variables.

Song, Romilly, and Liu (2000) measure the outbound tourism - total or per capita tourism visits (expenditure) from the origin country. For this purpose they employ an ECM framework and incorporate the income in origin country, the relative price, the substitute price, and the preference index which reflects psychological, social, and cultural influences on tourists' decisions. The comparison with a naïve model, as well as with AR(1), ARMA, and VAR demonstrates the superiority of the proposed approach. Shan & Wilson (2001) feed similar variables (tourist arrivals, income in origin country, relative price, exports, imports, and exchange rate) into a VAR model, but this time to estimate tourist arrivals.

Rosselló-Nadal (2001) forecasts tourist arrivals using ADLM with leading indicators, including exchange rates and economic indicators, and then compares its performance with an ARIMA model.

Song, Witt & Jensen (2003) embark on causal forecasting of the expenditure-weighted number of nights spent by tourists using two ECMs - one based on the WB approach and another one on the JML. The set of explanatory variables consists of the real private consumption expenditure per capita in origin country, the real cost of living for tourists (adjusted by the exchange rate), the prices in substitute destinations, the travel cost (airfare), and a number of dummy variables (oil crisis, Gulf War, German unification, and Chernobyl/U.S. bombing of Libya). The alternative models include static regression, ADLM, unrestricted VAR, TVP, and two univariate models (ARIMA and no-change model). The best performing models are found to be the TVP (for 1-year-ahead and 2—year-ahead forecasts) and the static model (2-, 3 – and 4- years-ahead forecasting horizons). The inferior performance of the VAR model may be attributed to the absence of any distinction between endogenous and exogenous variables. Witt, Song, and Louvieris (2003) use the same variables, but adds more data. The expanded dataset enhances the reliability of forecast error measurements. According to the new results, the VAR model outperforms its counterparts for 2- and 3-years-ahead forecasts, while the TVP and the ADLM are more robust for 1-year-ahead forecasts. Another contribution of this study is the statistical testing of directional change forecasting performance and unbiasedness. Song, Witt & Li (2003) base their analysis on a similar set of explanatory variables (they also consider trade volume and habit persistence). Those variables are fed into the following causal models: Reduced ADLM, Wickens–Breusch ECM, and Johansen maximum likelihood ECM. The alternative approaches include a Naïve model, ARIMA, and TVP. ARIMA seems to generate the most modest forecasts and the JML-ECM the most optimistic ones. In addition, habit persistence is found to be the most critical factor.

Nadal, Font, & Rossello (2004) develop an ECM model to describe the intra-year variation in tourist arrivals through the Gini-coefficient. The explanatory variables comprise income in origin country and exchange rates. The study of Dritsakis (2004) is based on a VAR framework and confirms that there is significant causality between domestic economic growth and tourism earnings.

Wong, Song, & Chon (2006) demonstrate that Bayesian VAR (BVAR) models overcome the weaknesses of VAR models for the prediction of tourism demand. The explanatory variables used include the income level in origin countries, the relative cost of living and a weighted average of the consumer price index of the major substitute destinations.

Kasimati (2011) employs a trivariate model consisting of exchange rates, the GDP and tourist arrivals. The ultimate goal is to explore the causal relationship between economic growth and demand for tourism in the context of VAR/VECM models. Unlike similar studies, the results do not support the existence of such a relationship.

Georgantopoulos (2012) adopts a similar modelling framework, but with different variables. In this case the underlying variables involve tourism expenditure, tourism growth and real output (RGDP) with real effective exchange rate (REER). In parallel, he develops another model which considers the following inputs: Leisure travel and tourism spending (LTS), business travel and tourism spending (BTS), RGDP and REER. Overall, the results show that all variables are co-integrated.

Chatziantoniou, Degiannakis, Eeckels, & Filis (2016) develop a SARMAX model incorporating the following macroeconomic indicators: Industrial production, CPI, CC index, BC index, Economic PUI, and the consumer price differentials (CPDs). The comparison with a SARIMA model show that the addition of exogenous variables improves the forecasting ability of this framework. However, none of these variables appears to be critical for all origin and destination combinations of this study.

Zhu, Lim, Xie, & Wu (2016) adopt a different approach to forecast tourist flows. They use a Frank copula model which draws on linkages between tourist flows and a set of variables (i.e. Income, relative price, and transportation cost). This model is compared with product copula and ARDL-ECM; The results demonstrate its superiority.

Table 2 lists the reviewed publications on causal forecasting models, presenting the model types used, the dependent and independent variables, the time period on which the forecasting is made, the models compared and the main outcome. For selected references, their contribution is also given at the end of Table 2.

*Table 2. Publications based on causal forecasting models*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reference | Model type | Dependent variable(s) | Independent variable(s) | Data | Comparison with | Outcome |
| Witt & Witt (1992) | ADLM | France, Germany, UK, and US tourist arrivals | Income in origin country, prices, exchange rate, travel cost (airfare), travel cost to substitute destinations, currency restrictions, dummy variables | Annual -  1965 - 1980 | Naive 1,2 AR, Exponential Smoothing, ARIMA Trends |  |
| Syriopoulos (1995) | CI/ECM | Mediterranean countries tourism receipts/expenditure | Y/P C RC ER D DT(ST) income in origin country, relative tourism price, exchange rate, dummy variable, deterministic (linear) trend (stochastic trend) | Annual - 1962 - 1987 | - |  |
| Song, Romilly, and Liu (2000) | ECM | UK outbound tourism - total or per capita tourism visits (expenditure) from the origin country | Income in origin country, relative price, substitute price, the preference index (incorporates social, cultural and psychological influences on tourists' decisions), dummy variable | Annual - 1965 - 1994 | Naive, AR(1), ARMA, VAR | ECM model has the best overall forecasting performance |
| Shan & Wilson (2001) | VAR | China tourist arrivals | Tourist arrivals, income in origin country, relative price, exports, imports, exchange rate | Montlhy - 1/1987 - 1/1998 | - |  |
| Rosselló-Nadal (2001) | ADLM (Leading indicator) | Balearic Islands tourist arrivals | Exchange rate, Other economic indicators | Monthly - 1/1975 - 12/1999 | Naïve ARIMA |  |
| Song, Witt & Jensen (2003) | ECM - Two ECMs, one based on the WB approach and the other on the JML approach | Denmark expenditure-weighted number of nights spent by tourists | Real private consumption expenditure per capita in origin country, the real cost of living for tourists (adjusted by the exchange rate), prices in substitute destinations, travel cost (airfare), dummy variables (oil crisis, Gulf War, German unification, Chernobyl/U.S. bombing of Libya | Annual - 1969 - 1997 | Static regression, ADLM, unrestricted VAR, TVP, and two univariate time-series models (ARIMA and no-change model) | The TVP model generates the most accurate 1-year-ahead forecasts. The TVP and static models generate the most accurate 2-years-ahead forecasts. For 3- and 4-years-ahead forecasts the static model is ranked first. |
| Witt, Song, and Louvieris (2003) | CI/ECM | Denmark expenditure-weighted number of nights spent by tourists | Real private consumption expenditure per capita, the real cost of living for tourists (adjusted by the exchange rate), prices in substitute destinations, travel cost (airfare), dummy variables (oil crisis, Gulf War, German unification, Chernobyl/U.S. bombing of Libya) | Annual - 1969 - 1999 | Static Regression, ADLM,Unrestricted VAR, the Time-Varying Parameter (TVP) approach - The univariate timeseries models comprise an ARIMA model and the no-change model | For both 2- and 3-years-ahead forecasting horizons,the VAR model generates the most accurate,unbiased forecasts. For a 1-year-ahead forecasting horizon,the TVP model and the reduced ADLM generate the most accurate,clearly unbiased forecasts |
| Song, Witt & Li (2003) (1) | Reduced ADLM, Wickens–Breusch ECM, and Johansen maximum likelihood ECM | Thailand tourist arrivals | Income in origin country, relative tourism price, substitute tourism price, trade volume, habit persistence, and dummy variables (two oil crises, Visit Thailand Year’ campaign, Asian financial crisis, Seoul Olympics, and student demonstrations in Korea in 1980) | Annual 1963/1968 - 2000 | ARIMA, Naive, TVP | ARIMA produces the most modest prediction (with the exception of Singapore), while the JML-ECM generates the most optimistic forecasts. Habit persistence is the most important factor. |
| Nadal, Font, & Rossello (2004) (2) | ECM | Intra-year variation in arrivals to the Balearic Islands | Y/P ER RC income in origin country, exchange rate | Monthly/Annual - 1982 - 2001 | - |  |
| Dritsakis (2004) | VAR | Long-run economic growth of Greece (real gross domestic product & real effective exchange rate) | International tourism earnings | 1960-2000 | - | Significant causality |
| Wong, Song, & Chon (2006) (3) | Bayesian VAR (BVAR) | Demand for Hong Kong tourism by residents from six major origin countries | Income level of origin country, the relative cost of living, the Substitute Price variable | Annual - 1973 - 1996 | unrestricted VAR models | Univariate Bayesian VAR (BVAR) the best performing model |
| Kasimati (2011) | VAR/VECM - Trivariate model - Granger causality | Economic growth of Greece (GDP, employment) | International tourist arrivals | Annual data - 1960-2010 | - | No direct causality |
| Georgantopoulos (2012) | VAR/VECM - Trivariate model | Tourism expenditure and growth in Greece | Total tourism expenditure (TE) and real output (RGDP) with real effective exchange rate (REER) | Annual data - 1988-2012 | A second model treated leisure travel and tourism spending (LTS), business travel and tourism spending (BTS), RGDP and REER as separate inputs | All variables return to their long-term equilibrium |
| Chatziantoniou, Degiannakis, Eeckels, & Filis (2016) | SARMAX (with exogenous macroeconomic variables) | Tourist arrivals to Greece from seven key origin countries(Aggregate data) - Disaggregated data by origin country | Macroeconomic indicators: Industrial production, consumer price index (CPI), CC index, BC index, Economic PUI, and the consumer price differentials (CPDs) between Greece and the origin countries (estimated using the Greek CPI) - | Monthly data | SARIMA | SARMAX outperforms SARIMA - No single variable improves the forecasting accuracy for all cases - The CC index is critical for Canada and the UK - The price level for France and Germany - Income for Italy and Spain - Economic policy uncertainty for the USA. |
| Zhu, Lim, Xie, & Wu (2016) (4) | Frank copula-based model | Tourist flows to Singapore from 6 countries | Income, relative price, and transportation cost - Exponential smoothing is used to predict them | Monthly data - 1995 - 2013 | Product copula, ARDL-ECM | Frank model performs better |

Song, Witt & Li (2003) (1) The relatively poor forecasting performance of the VAR model for all forecasting horizons other than 4 years ahead may have arisen because it is the only causal model included in the forecasting comparison which generates ex ante forecasts. Alternatively, however, the low ranking may suggest that the distinction between endogenous and exogenous variables in tourism demand forecasting models is fairly clear, or at least that allowing for the distinction not to be clear does not result in more accurate forecasts except perhaps in the longer term.

Nadal, Font, & Rossello (2004) (2) The rigorous statistical testing of directional change forecasting performance and unbiasedness presented in the current study is new to the tourism forecasting literature

Wong, Song, & Chon (2006) (3) The reason why Bayesian approach tend to improve the forecasting performance of the unrestricted VAR model may be attributed to the introduction of the Bayesian priors. These priors serve to improve the estimation efficiency through increasing the degrees of freedom and controlling for over-parameterization associated with the unrestricted VAR models.

Zhu, Lim, Xie, & Wu (2016) (4) Unlike prior tourism research, we take into account the dependence relations among the different tourist flows via copula.

**Machine Learning**

Techniques based on Machine Learning and Artificial Neural Networks have gained traction over the past few years and constitute a relatively new trend in tourism forecasting.

Atsalakis, Chnarogiannaki, and Zopounidis (2014) forecast tourist arrivals using Adaptive Neuro-Fuzzy Inference System (ANFIS). They also fit Autoregressive (AR) & autoregressive moving average (ARMA) models. The comparison shows that the ANFIS outperform the alternative approaches.

Claveria, Monte, and Torra (2015) compares the performance between Radial basis function neural networks and multi-layer perceptron and Elman recursive neural networks. According to the results, the first approach forecasts tourist arrivals more accurately. In addition, Claveria, Monte, & Torra (2016) forecast tourist arrivals through an extension of the Gaussian process regression (GPR) model to a multiple-input and multiple-output framework that take account of the cross-interactions. The authors show that the proposed model outperforms the standard neural network, which is use as a benchmark.

Kummong and Supratid (2016) uses Discrete Wavelet Decomposition (DWD)-NARX - Combination of DWD and nonlinear autoregressive neural network with exogenous input (NARX). This methodology appears to generate better predictions of tourist arrivals in the short-run than alternative neural network approaches, such as GRNN, Elman’s RNN and BPNN.

Yu, Wang, Gao, and Tang (2017) adopt seasonal trend autoregressive integrated moving averages with Dendritic Neural Network (SA-D model) with the aim of forecasting the number of inbound tourists. The results suggest that this model performs better that the combined SARIMA and DNN.

Table 3 lists the reviewed publications on Machine Learning and Artificial Neural Networks models, presenting the model types used, the dependent and independent variables, the time period on which the forecasting is made, the models compared and the main outcome. For selected references, their contribution is also given at the end of the Table 3.

*Table 3. Publications based on Machine Learning and Artificial Neural Networks*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Reference | Model type | Dependent variable(s) | | Independent variable(s) | | | Data | Comparison with | | Outcome | |
| Atsalakis, Chnarogiannaki, & Zopounidis (2014) | Adaptive Neuro-Fuzzy Inference System (ANFIS) | | Tourist arrivals to Greece by air, sea, and land transport | |  | Monthly – 1/1996 – 11/2011 | | | Autoregressive (AR) & autoregressive moving average (ARMA) | | The ANFIS outperformed the other models |
| Claveria, Monte, & Torra (2015) (1) | Radial basis function neural networks | | Tourist arrivals to Catalonia | | Correlations in the evolution of inbound international tourism demand | Monthly - 1/2001 -  7/2012 | | | Multi-layer perceptron and Elman recursive neural networks | | Radial basis function neural networks performs better |
| Claveria, Torra, & Monte (2016) (2) | Machine learning techniques - Support Vector Regression (SVR) | | International tourist arrivals to key Spanish regions | |  | Monthly - 1/1999 to 3/2014 | | | Neural Network (NN), ARMA | | SVR with a Gaussian radial basis function kernel performs better than the other models |
| Claveria, Monte, & Torra (2016) (3) | Extension of the Gaussian process regression model for multiple-input multiple-output forecasting | | Tourist arrivals to all Spanish regions | | Cross-correlations in tourist arrivals to all markets | Monthly data - 1/1999 - 3/2014 | | | Standard neural network | | Extension of Gaussian outperforms the benchmark |
| Kummong & Supratid (2016) (4) | Discrete Wavelet Decomposition (DWD)-NARX - Combination of DWD and nonlinear autoregressive neural network with exogenous input (NARX) | | Thailand tourist arrivals | |  | Relative price with respect to 6 top-ranked tourist countries (exogenous) | | | NARX and other neural network approaches (GRNN, Elman’s RNN and BPNN) | | DWD-NARX more suitable for short-term projections |
| Yu, Wang, Gao, Tang (2017) (5) | Seasonal trend autoregressive integrated moving averages with Dendritic Neural Network (SA-D model) - Combined SARIMA and DNN | | Inbound tourists - Japan | |  | 1/2009 - 12/2015 | | | Dendritic Neural Network (DNN) | | SA-D performs better |

Claveria, Monte, & Torra (2015) (1) A way of using the common trends in tourist arrivals from different visitor markets and assessing its performance.

Claveria, Torra, & Monte (2016) (3) The main contribution of this study to the previous literature on tourism demand forecasting is the evaluation of the performance of several ML techniques by means of an iterated multi-step ahead forecasting comparison.

Claveria, Monte, & Torra (2016) (3) Incorporating the connections between different markets in the modelling process may prove to be useful in refining predictions at a regional level.

Kummong & Supratid (2016) (4) (DWD)-NARX deals with non-stationarity and overcomes overfitting.

Yu, Wang, Gao, Tang (2017) (5) The study is based on neuron model with dendritic nonlinearity model and it theoretically strengthens the assumption that a neural network model performs better than linear models when forecasting nonlinear variables. This study mixed linear and nonlinear models and opens the door for further combinations.

**Miscellaneous/ Alternative modelling approaches**

Data limitations are a common barrier to the robust performance of most tourist forecasting techniques. This has compelled some researches to seek new sources of data. Thus, some of them have based their analysis on web search data. For example, Peng, G., Liu, Y., Wang, J., and Gu, J. (2016) estimate the number of visitors fitting web search query data into a HE-TDC screening model. The comparison with classical regression analysis reveals that this new methodology can improve the forecasting accuracy. Li, Pan, Law, and Huang (2017) use large search trend data and construct a Composite search index on the basis of a Generalized Dynamic Factor Model (GDFM) to forecast tourism demand (including tourist numbers and hotel occupancy). This model outperforms both a time-series model and a model based on Principal Component Analysis (PCA).

Some other authors have taken a different path towards forecasting tourist arrivals. They have departed from the typical techniques in the tourism forecasting literature and test novel approaches. For example, Chinnakum and Boonyasana (2016) develop a Kink AR-GARCH models to predict tourist arrivals. This model appears to improve the forecasting ability, considering that the AR(m)- GARCH(p,q) which is used for comparison purposes performs worse. In another study Wan, Song, and Ko (2016) embark on density forecasting to predict the growth rate of tourist arrivals. The explanatory variables include the GDP and the exchange rate-adjusted consumer price index.

Table 4 lists the reviewed publications on alternative methodologies than those mentioned before. It has the same structure with Tables 1,2 and 3.

*Table 4. Publications based on Miscellaneous/ Alternative modelling approaches*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reference | Model type | Dependent variable(s) | Independent variable(s) | Data | Comparison with | Outcome |
| Peng, G., Liu, Y., Wang, J., & Gu, J. (2016) | HE-TDC screening method (A new method that combines Hurst Exponent (HE) and Time Difference Correlation (TDC) analysis) | Volume of tourism visitors in Jiuzhai Valley |  | Web search query data | Regression analysis | The HE-TDC method improves the predictive ability. |
| Li, Pan, Law, & Huang (2017) | Composite search index in the framework of a Generalized Dynamic Factor Model (GDFM) | Tourism demand in Beijing, including tourist numbers and hotel occupancy |  | Large search trend data | A traditional time-series model and a model with an index created by Principal Component Analysis (PCA) | GDFM improves the forecasting accuracy |
| Chinnakum & Boonyasana (2016) | Kink AR-GARCH | Tourist arrivals to Thailand from East Asia | - | Monthly data - 1/1991 - 2/2016 | AR(m)- GARCH(p,q) | Kink AR(1)-GARCH(1,1) performs better |
| Wan, Song & Ko (2016) (1) | Density forecasting - ARDL (p,q,r) | Y-O-Y growth rate of tourist arrivals to Hong Kong | GDP (growth) and Exchange rate-adjusted consumer price index (infl) | Quarterly data - 1996Q1 - 2015Q1 | - |  |

Wan, Song & Ko (2016) (1) This research attempts to alleviate this risk by introducing density forecast for  
tourism demand that accommodates both mean forecast and its uncertainty in the analysis. Interval forecast is an immediate response to the above criticism on point forecast as it specifies the probability that the actual outcome will fall within a stated interval.

**Areas of focus**

Most of the previously mentioned studies focus on Asian and European countries. Specifically, Shan and Wilson (2001), Wong et al. (2006), Cho (2001), ShuiKi et al. (2013), Wan et al. (2016), Wan et al. (2016), Peng et al. (2016), Li et al. (2017) concentrate on Hong Kong and China, while Song and Witt (2003), Kim and Uysal(1998) on Korea. Furthermore, Song et al. (2003), Zhu et al. (2016), Kummong and Supratid (2016), Chinnakum and Boonyasana (2016), Yu et al. (2017) consider other Asian destinations.

Several other studies refer to Mediterranean and European countries - especially Spain Witt and Witt (1992), Syriopoulos (1995), Song et al. (2000), Rosselló & Nadal (2001), Song et al., (2003), Witt et al. (2003), Nadal et al. (2004), Akal (2004), Baldigara and Mamula (2015), Claveria et al., (2015), Claveria et al. (2016), Claveria et al.(2016), and Hassani et al. (2017).

Finally, there is another growing strand in the literature that focuses on Greece and the prediction of tourism demand there Dritsakis (2004), Kasimati (2011), Gounopoulos et al., (2012), Georgantopoulos (2012), Atsalakis et al., (2014), and Chatziantoniou et al., (2016). This is consistent with the objectives of the present study, which pertain to the specification of suitable forecasting models for Greek tourism.

Table 5 presents the selected papers for review in terms of the region/country they examine.

*Table 5. Publications presented by region / country*

|  |  |  |
| --- | --- | --- |
| Region / Country | | References |
| Europe | Greece | Dritsakis (2004), Kasimati (2011), Gounopoulos et al., (2012), Georgantopoulos (2012), Atsalakis et al., (2014), and Chatziantoniou et al., (2016). |
|  | Spain | Witt and Witt (1992), Syriopoulos (1995), Song et al.(2000), Rosselló & Nadal (2001), Song et al., (2003), Witt et al. (2003), Nadal et al. (2004), Akal (2004), Baldigara and Mamula (2015), Claveria et al., (2015), Claveria et al. (2016), Claveria et al. (2016), and Hassani et al. (2017). |
| Asia | China, Hong Kong | Shan and Wilson (2001), Wong et al. (2006), Cho (2001), ShuiKi et al. (2013), Wan et al. (2016), Wan et al. (2016), Peng et al. (2016), Li et al. (2017) |
|  | Korea | Song and Witt (2003), Kim and Uysal (1998) |
|  | Other | Song et al. (2003), Zhu et al. (2016), Kummong and Supratid (2016), Chinnakum and Boonyasana (2016), Yu et al. (2017) |

## **References**

|  |
| --- |
| Akal, M. (2004). Forecasting Turkey's tourism revenues by ARMAX model. Tourism Management, 25(5), 565-580. |
| Atsalakis, G., Chnarogiannaki, E., & Zopounidis, C. (2014). Tourism demand forecasting based on a neuro-fuzzy model. Hospitality, Travel, and Tourism: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications, 106. |
| Baldigara, T., & Mamula, M. (2015). Modelling international tourism demand using seasonal ARIMA Models. Tourism and Hospitality Management, 21(1), 19-31 |
| Chatziantoniou, I., Degiannakis, S., Eeckels, B., & Filis, G. (2016). Forecasting tourist arrivals using origin country macroeconomics. Applied Economics, 48(27), 2571-2585. |
| Chinnakum, W., & Boonyasana, P. (2016). Forecasting International Tourism Demand in Thailand. Thai Journal of Mathematics, 231-244. |
| Cho, V. (2001). Tourism forecasting and its relationship with leading economic indicators. Journal of Hospitality & Tourism Research, 25(4), 399-420. |
| Claveria, O., Monte, E., & Torra, S. (2015). Common trends in international tourism demand: Are they useful to improve tourism predictions?. Tourism Management Perspectives, 16, 116-122. |
| Claveria, O., Monte, E., & Torra, S. (2016). Combination forecasts of tourism demand with machine learning models. Applied Economics Letters, 23(6), 428-431. |
| Claveria, O., Monte, E., & Torra, S. (2016). Modelling cross-dependencies between Spain’s regional tourism markets with an extension of the Gaussian process regression model. SERIEs, 7(3), 341-357. |
| Dritsakis, N. (2004). Cointegration analysis of German and British tourism demand for Greece. Tourism management, 25(1), 111-119. |
| Georgantopoulos, A. G. (2012). Forecasting tourism expenditure and growth: A VAR/VECM analysis for Greece at both aggregated and disaggregated levels. International Research Journal of Finance and Economics, 96(105), 155-167. |
| Gounopoulos, D., Petmezas, D., & Santamaria, D. (2012). Forecasting tourist arrivals in Greece and the impact of macroeconomic shocks from the countries of tourists’ origin. Annals of Tourism Research, 39(2), 641-666. |
| Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., & Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. Annals of Tourism Research, 63, 112-127. |
| Kasimati, E. (2011). Economic impact of tourism on Greece’s economy: Cointegration and causality analysis. International research journal of Finance and Economics, 79(5), 79-85. |
| Kummong, R., & Supratid, S. (2016). Thailand tourism forecasting based on a hybrid of discrete wavelet decomposition and NARX neural network. Industrial Management & Data Systems, 116(6), 1242-1258. |
| Kim, Y., & Uysal, M. (1998). Time-dependent analysis for international hotel demand in Seoul. Tourism Economics, 4(3), 253-263 |
| Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. Tourism Management, 59, 57-66. |
| Nadal, J. R., Font, A. R., & Rossello, A. S. (2004). The economic determinants of seasonal patterns. Annals of Tourism Research, 31(3), 697-711 |
| Peng, G., Liu, Y., Wang, J., & Gu, J. (2016). Analysis of the prediction capability of web search data based on the HE-TDC method‒prediction of the volume of daily tourism visitors. Journal of Systems Science and Systems Engineering, 1-20. |
| Rosselló-Nadal, J. (2001). Forecasting turning points in international visitor arrivals in the Balearic Islands. Tourism Economics, 7(4), 365-380. |
| Shan, J., & Wilson, K. (2001). Causality between trade and tourism: empirical evidence from China. Applied Economics Letters, 8(4), 279-283. |
| ShuiKi, W., ShinHuei, W., & ChiKeung, W. (2013). Aggregate vs. disaggregate forecast: case of Hong Kong. Annals of Tourism Research, 42, 434-438. |
| Song, H., Romilly, P., & Liu, X. (2000). An empirical study of outbound tourism demand in the UK. Applied economics, 32(5), 611-624. |
| Song, H., Witt, S. F., & Jensen, T. C. (2003). Tourism forecasting: accuracy of alternative econometric models. International Journal of Forecasting, 19(1), 123-141. |
| Song, H., Witt, S. F., & Li, G. (2003). Modelling and forecasting the demand for Thai tourism. Tourism Economics, 9(4), 363-387. |
| Syriopoulos, T. C. (1995). A dynamic model of demand for Mediterranean tourism. International review of applied economics, 9(3), 318-336. |
| Wan, S. K., Song, H., & Ko, D. (2016). Density forecasting for tourism demand. Annals of Tourism Research, 60, 27-30. |
| Witt, S. F., Song, H., & Louvieris, P. (2003). Statistical testing in forecasting model selection. Journal of Travel Research, 42(2), 151-158. |
| Witt, S. F., & Witt, C. A. (1992). Modeling and forecasting demand in tourism. Academic Press Ltd.. |
| Wong, K. K., Song, H., & Chon, K. S. (2006). Bayesian models for tourism demand forecasting. Tourism Management, 27(5), 773-780. |
| Yu, Y., Wang, Y., Gao, S., & Tang, Z. (2017). Statistical Modeling and Prediction for Tourism Economy Using Dendritic Neural Network. Computational intelligence and neuroscience, 2017. |
| Zhu, L., Lim, C., Xie, W., & Wu, Y. (2016). Modelling tourist flow association for tourism demand forecasting. Current Issues in Tourism, 1-15. |