ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

MSc Applied Economics

Modelling and Forecasting the outbound tourist arrivals of Cyprus

Maragkopoulos Panagiotis

Contents

Abstract

The objective of this paper is to make accurate forecasts about tourist arrivals in Cyprus. It is undeniable that the sector of tourism and hospitality constitutes one of the most significant for the economic growth of the island. As a consequence, the ability to predict the future tourist demand is extremely useful both for the government and the stakeholders and can help in further development of this sector. Accurate forecasts provide valuable aid for the development of marketing and tourism strategies, and the pricing policies. In this investigation we use a variety of econometric models in order to make three years predictions about tourist arrivals in Cyprus and compare the accuracy of forecasts of these models through widely used forecast accuracy measures.

1.Introduction

Tourism has a large impact in world economy, creating direct and indirect benefits to all the industries within a region. It contributes to a country's economy in terms of employment, affects the balance of payment and investment and generates income and profit (see Ali, Ciaschini, Pretaroli, Severini and Socci (2014). According to the World Travel and Tourism Council (WTTC), in the year 2019 the Travel and Tourism sector has created worldwide an 8.9US $ contribution to the world's GDP, participated to the global GDP with 10.3% and created 330 million jobs worldwide.

Tourism industry is one of the major contributors to foreign exchange earnings for Cyprus as it occupies a dominant position in the economy. Moreover, it significantly affects Cyprus culture and its multicultural development throughout the decades. Cyprus has been a full member of the World Tourism Organization since 1975. More than 4 million tourists visit Cyprus every year and as a result the island is the 40th most popular destination in the world.

Tourism contributes highly to GDP, increasing the employment rate, sources of revenue for local people, private sector and public sectors. In addition, tourism constitutes significant part of government budget and strategy because the revenues of taxes are used to finance public investment plans. The tourism sector is dynamically developing in Cyprus and the contribution of Travel and Tourism to national GDP and employment is very high. In 2018, the contribution of travel and tourism to GDP (% of GDP) for Cyprus was 21.9 %. Before contribution of travel and tourism to GDP of Cyprus started to increase to reach a level of 21.9 % in 2018, it went through a trough reaching a low of 14.3 % in 2010. In addition, more than 85 thousand people are employed in this economic sector. Moreover, by 2025, the relative contribution of the tourism sector to overall economic activity is anticipated to be 25.5%. The statistical evidence reveals the significant contribution of Tourism in Cyprus economy. The ideal climate in combination with the developed tourism infrastructure makes Cyprus a very popular summer destination. Although Cyprus attracts tourists all year round, it is obvious that in summer season the tourist demand increases rapidly. For this reason, notable role in our analysis possesses the term of seasonality.

In this empirical research we desire to find the most appropriate and accurate model for forecasting of tourist arrivals in Cyprus. For this reason, we use the available data and make predictions in sample in order to evaluate the accuracy of different models for our case. The structure of our paper is organized as follows: In section 1 we make a brief literature review, in section 2 we describe the methodologies and the dataset that are used, in section 3 we present the empirical results of our investigation and we make some comments and the last section summarizes this paper.

2.Literature Review

With the importance of the tourism industry in Cyprus' economy, and the whole world in general, rising, the significance of tourism forecasts becomes more obvious as the time goes by. The benefits of accurate forecasts in tourism have been under a lot of study and as a result, the forecasting literature is very rich (see Archer 1987, Morley 1991, Frechtling 1996 and Smith 2014). Accurate forecasts are very valuable to both the public and the private sectors in order to avoid shortages or surpluses in goods and services and thus, it's significance in tourism planning, cannot be denied. Bull (1995) indicates that forecasting tourist numbers is very helpful in assessing the impact that tourism will make on the resources of an economy. Tourism investment should be based on professional business planning, long-term operating viability and an achievable vision of the industry's future (Lundberg, Krishnamoorthy and Stavenga,1995). One of the major needs of tourism industry is to reduce the risk of poor decisions. One way to achieve this is by better discerning the future events (see Smith 2014).

The models used in our analysis are proposed by the articles focusing on forecasting tourist arrivals. The most widely used procedures in non-causal time series forecasting are the autoregressive integrated moving average (ARIMA) models (Goh and Law, 2002) and the exponential smoothing (ES) models (Cho, 2003). Cho (2001). Song, Li, Witt, & Athanasopoulos (2011) developed the dynamic modelling through the integration of the time-varying-parameter (TVP) technique and the causal structural time series model. Claveria and Torra (2014) evaluate the forecasting performance of different models on overnight stays and tourist arrivals in Catalonia, amongst them, the SETAR model. Liang (2014) employs the SARIMA–GARCH model to analyse and predict tourism demand in Taiwan and makes a comparison with other models, such as ES and HW. Andrea Saayman and Ilse Botha (2016) used non-linear methods such as unobserved components model (BSM), smooth transition autoregressive models (STAR) and singular spectrum analysis to predict the tourist arrivals in South Africa. Comparing the results of non-linear models with the results of SARIMA and Naïve model concluded that the non-linear models presented better forecasting performance. BSM model is also applied by Du Preez and Witt (2003) and Turnen and Witt (2001) for time-series forecasting. Mamula (2015) examines the forecasting tourism demand in Croatia by incorporating seasonal dummies in Linear Regression model to capture the seasonality effect. Gunter, U., Önder, I., & Gindl, S. (2019) incorporated in their investigation Google Trends in order to test whether forecast models with LIKES and/or with Google Trends deliver more accurate forecasts. To capture the dynamics in the data, the autoregressive distributed lag (ADL) model class is employed.

3. Methodology

We use three simple models as benchmark in order to have direct comparisons with other more complicate models.

3.1.1 Naïve Model

The Naïve model assumes that tourism arrivals follow a random walk, and trends and turning points can therefore not be predicted. We make the assumption that the tourist arrivals in the current period (F_t) are equal to arrivals in the previous period (Y_{t-1}).

$F_t=Y_{t-1}$ (3.1)

3.1.2 Linear Regression (LR)

Having as dependent variable the tourist arrivals and as independent variables a time trend, a constant and eleven seasonal dummies, we conduct a forecast of tourist arrivals through a simple linear regression with OLS method. The model that we estimate is:

$\text{Arrivals}= c + a_1 \text{ trend}+ b_1d_1+ b_2d_2+\ldots b_{11}d_{11} + e_t$ (3.2)

Where c is a constant, trend is a time trend and d_1, d_2... d_{11} are the seasonal dummies. We include eleven seasonal dummies instead of twelve in order to avoid the trap of dummies variable and multicollinearity problems.

3.1.3 Moving Average (MA)

The moving average of order K evaluated at time t is denoted by MA(t/K):

$$MA(t/K)=\frac{Y_t+Y_{t-1}+\ldots+Y_{t-K+1}}{K} \quad (3.3)$$

At each point in time t we remove the oldest observation and we add a new one. In our case we use three types of Moving Average models in order to produce forecasts for tourist arrivals in Cyprus. Moving average models of 3, 6 and 12 observations are used. The formula that we use to forecast the time series at time t+1 is:

i) MA(3): $F_{t+1}=\frac{1}{3}*\sum_{(i=t-3+1)}^{t} Y_i$

ii) MA(6): $F_{t+1}=\frac{1}{6}*\sum_{(i=t-6+1)}^{t} Y_i$

iii) MA(12): $F_{t+1}=\frac{1}{12}*\sum_{(i=t-12+1)}^{t} Y_i$

3.1.4 Autoregressive Model (AR)

A simple AR(1) model is used:

$$Y_t=\delta+\varphi Y_{t-1}+\varepsilon_t \quad (3.4)$$

3.2 Exponential Smoothing

All exponential smoothing methods also average exponentially the data. The exponential averaging of data means that, the more recent observations receive more weight than the less recent ones. There are many different methods of exponential smoothing: the Simple Exponential Smoothing, the Double Exponential Smoothing and the three methods of Holt Winters Exponential Smoothing (Without Seasonal component, with Additive and Multiplicative Seasonal component). The main difference among the various exponential smoothing methods is the way they treat the trend and seasonality. In this paper we use four out of five methods were referred above.

3.2.1 Single Exponential Smoothing

The single exponential smoothing is an appropriate method for forecasting time series without trend and seasonality.

The forecasting formula is the basic equation

$$S_{t+1}=aY_t+(1-a)S_t, 0<a\leq 1, t>0$$

This can be written as:

$$S_{t+1}=St+a\,e_t$$

where e_t is the forecast error (actual - forecast) for period t. In other words, the new forecast is the old one plus an adjustment for the error that occurred in the last forecast.

3.2.2 Double Exponential Smoothing

This method applies the single smoothing method twice (using the same parameter) and is appropriate for series with a linear trend. Double smoothing of a series Y is defined by the recursions:

$$S_{t+1}=aY_t+(1-a)S_{t-1}$$

$$D_{t+1}=aS_t+(1-a)D_{t-1}$$

where S is the single smoothed series and D is the double smoothed series and $0< a\leq 1$

Forecasts from double smoothing are computed as

$$F_{t+k}=\left(2S_t-D_t+\frac{a}{1-a}(S_t-D_t)\right)k$$

3.2.3 Holt-Winter Exponential Smoothing (Additive)

This model is considered appropriate for time series depicting a linear time trend and additive seasonal variation. (Holt, 1957; Winters, 1960). This method employs a triple exponential smoothing framework comprising level, trend and seasonality equations. These are specified as following:

The forecast formula is:

$$F_{t+1}=L_t+T_t+S_{t-m+1}$$

The observed error is:

$$e_t=Y_t-L_{t-1}-T_{t-1}+S_{t-m}$$

The updating relationships are:

$$L_t=L_{t-1}+T_{t-1}+\alpha(Y_t-L_{t-1}-S_{t-m})$$

$$T_t=T_{t-1}+\beta(L_t-L_{t-1}-T_{t-1})$$

$$S_t=S_{t-m}+\gamma(Y_t-L_{t-1}-T_{t-1}-S_{t-m})$$

3.2.4 Holt-Winter Exponential Smoothing (Multiplicative)

This model is considered appropriate for time series depicting a linear time trend and multiplicative seasonal variation (Holt, 1957; Winters, 1960). The forecast function includes the local trend and seasonal components as before, but the seasonal effect is now multiplicative. This method employs a triple exponential smoothing framework comprising level, trend and seasonality equations. These are specified as under:

The forecast function:

$$F_{t+1}=\left(L_t+hT_t\right)S_{t-m+h}$$

The error correction updating relationships are:

$$L_t=L_{t-1}+T_{t-1}+\alpha\left(e^t/S_{t-1}\right)$$

$$T_t=T_{t-1}+\alpha\beta\left(e_t/S_{t-1}\right)$$

$$S_t=S_{t-m}+\gamma\left(e_t/S_{t-1}\right)$$

3.3 ARIMA Models

3.3.1 Auto Regressive Integrated Moving Average (ARIMA)

ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting, and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

Univariate ARIMA models use only the information contained in the series itself. Therefore, these models are constructed as linear functions of past values of the series and/or previous random shocks. Forecasts are generated under the assumption that the past history could be translated into predictions for the future. The ARIMA model uses the fact that arrival of tourist's is a stochastic time series. This modelling regresses the dependent variable Y_t on p-lags of the dependent variable (Autoregressive) and q lags of the error term (Moving Average).

$$Y_t=\alpha_1 Y_{t-1}+\alpha_2 Y_{t-2}+\ldots+\alpha_p Y_{t-p}+\gamma_1 e_{t-1}+\gamma_2 e_{t-2}+\ldots+\gamma_p e_{t-p}+u_i$$

We follow the steps below in ARIMA modelling:

1) The Autocorrelation function (ACF) and the Partial Autocorrelation function are used to identify the model. These functions measure the statistical correlation within the time series data. One has to select the model whose theoretical ACF and PACF resembles the expected ACF and PACF of the time series data.

2) The method of Maximum Likelihood Estimation (MLE) or of the Modified Least Squares (MLS) is used to estimate the parameters of ARIMA models.

3) To identify the optimal ARIMA model different combinations of AR and MA are tested and we choose the model which minimizes the Akaike Information Criterion (AIC).

4) Diagnostic test in residuals of the fitted model.

5) Forecasting

3.3.2 Seasonal Auto Regressive Integrated Moving Average (SARIMA)

The seasonal ARIMA model incorporates both seasonal and non-seasonal factors in a multiplicative model. The non-seasonal autoregressive and moving average terms accounting for the correlation at lower lags are used in SARIMA model. Similarly, the seasonal autoregressive and moving average terms accounting for the correlation at seasonal lags are also contained in SARIMA. Seasonal ARIMA model is popular due to its ability to deal with both stationary and non-stationary series. In case of non-stationary time series, the values of the variable are taken in their first differences to estimate the model. The seasonal ARIMA model is usually represented by a multiplicative model in the form of SARIMA (p, d, q) (P, D, Q)$_s$ where p is the non-seasonal autoregressive order d is the non-seasonal differencing q is the non-seasonal moving average order, is the seasonal autoregressive order, D is the seasonal differentiation, Q is the seasonal moving average order; and is the time span of repeating seasonal pattern.

$$Y_t = \varphi_1 Y_{t-1} + \Phi_{12} Y_{t-12} + \varphi_1 \Phi_{12} X_{t-13} + \xi_t - \theta \xi_{t-1} - \Theta_{12} e_{t-12} - \theta_1 \Theta_{12} \xi_{t-13} + u_i$$

with φ_i and θ_i the non-seasonal parameters that are estimates, Φ_i and Θ_i the seasonal parameters, and ξ_t an uncorrelated random shock.

3.4.1 Self- exciting Threshold Autoregressive Model (SETAR)

SETAR models were initiated by Howell Tong in 1977 as an extension of AR models designed to handle changes in the model parameters by the threshold value and delay parameter. Tong and Lim stated that SETAR (1, K) is a linear AR model with order k. Serletis and Shahmoradi stated the following model as two regimes SETAR model.

$$Y_t = a_0 + a_1 Y(t-1) + \ldots + a_p Y(t-p) + \left(\beta_0 + \beta_1 Y(t-1) + \ldots + \beta_p Y(t-p)\right)\left|Y(t-d) \le \gamma\right| + \varepsilon_t \quad (2.4.1)$$

where p ≥ 1 = autoregressive order, d = delay parameter and γ = the threshold parameter.

3.4.2 Unobserved Component Model

The BSM decomposes the time series into three independent components, namely a trend, seasonal and irregular component (Harvey and Peters, 1990) and has the advantage that it can treat these components as stochastic. The BSM can be represented in the following state space form (SSF):

$Y_t = \mu_t + \gamma_t + e_t, e_t \;\; NID(0, H_t)$ (3.4.1)

$\mu_t = \mu_{t-1} + \beta_{t-1} + u_t, u_t \;\; NID(0, \sigma_u^2)$ (3.4.2)

$\beta_t = \beta_{t-1} + \delta_t, \delta_t \;\; NID(0, \sigma_\delta^2)$(3.4.3)

$\gamma_t = -\sum_{i=1}^{s-1} \gamma_{t-j} + \kappa_t, \kappa_t \;\; NID(0, \sigma_\kappa^2)$(3.4.4)

where Y_t is a univariate time series, decomposed into its unobservable components, including a trend component (μ_t), a seasonal component (γ_t) and an irregular component (ε_t).

Equations (3.4.2) and (3.4.3) specifies the stochastic trend while the β is the slope of trend. The seasonal component is defined in Equation 2.4.4) in such a stochastic way that the seasonal pattern is allowed to change over time, where s is the number of seasons per year. It is often preferable to express the stochastic seasonality in trigonometric form. The white noise disturbances of the trend and seasonal equations (Eqs. (3.4.2)–(3.4.4)) are independent and σ_u^2 , σ_δ^2 and σ_κ^2 are the corresponding variances.

3.5 GARCH Models

3.5.1 AR(1)- GARCH(1,1)

The GARCH model has the ability to model time-varying conditional variances. In our GARCH(1,1) model, we incorporate one lag of our dependent variable in the mean equation. For univariate series:

$y_t = \mu + y_{t-1} + e_t$ (3.5.1)

The equation (3.5.1) is a mean equation at a time, where μ is the conditional mean of the yt and et is the shock at time t. The equation of conditional variance is:

$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i * e_{t-i}^2 + \sum_{i=1}^{p} \gamma_i * \sigma_{t-i}^2$ (3.5.2)

Where $\alpha_0 > 0$ and $\alpha_i + \gamma_i < 1$, where a_i and γ_i are the coefficients of the parameters ARCH(e_{t-i}^2) and GARCH (σ_{t-i}^2) respectively.

3.5.2 AR(1)-EGARCH

Exponential-Garch model is suggested by Nelson (1991). The mean equation of AR(1)-EGARCH remains the same with AR(1)- EGARCH.

$y_t = \mu + y_{t-1} + e_t$ (3.5.3)

The specification for conditional variance is:

$$\log\left(\sigma_t^2\right) = \omega + \beta \log\left(\sigma_{t-1}^2\right) + \gamma\left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) + \left[\frac{|u_{t-1}|}{\sigma_{t-1}} - \frac{2}{\sqrt{\pi}}\right] \text{ (2.5.4)}$$

3.6 Google search engine and tourist arrivals prediction

Google Trends data show the relative volume of web searches relating to a particular keyword, which can be defined by the user. In tourism studies, destination names, hotels, flights, and other travel-related keywords are used to retrieve the Google Trends data applied in tourism demand forecast models. The predictive ability of Google Trends data has been investigated in various areas, while there are 65% predictable queries within the travel category.

3.6.1 OLS

Firstly we use a simple OLS regression of Tourist arrivals time series on a constant and the google searches data. The selected key phrase which we chose to incorporate in our forecasts is " hotel Cyprus". Using the estimation of the model below, we conduct the first forecast through the use of google trends.

d[Log(Arrivals)]= c + dD[Log(Googletr)]+e_t (3.6.1)

3.6.2 SARIMAX

Having two variables that present strong evidence of seasonality we select to estimate a Seasonal ARIMAX model. In this stage, we include the variable of google trends as exogenous in the estimation SARIMA model. In particular we estimate MSARIMA (multivariate SARIMA) or SARIMAX model. The methodology remains the same with the case of univariate models. Aided by the software we estimate all possible models and we select the one that minimizes the information criteria (Akaike).

3.6.3 ARDL

The autoregressive distributed lag (ADL) model class is employed in order to capture the dynamics in the data. The model that we estimate is:

$$TA_t = a_j \sum_{i=1}^{12} b_i \times TA_{t-i} + \sum_{i=0}^{12} d_i \times TREDS_{t-i} + e_t \text{(3.6.2)}$$

$TRENDS_{t-i}$ denotes the current and past monthly Google Trends web search index .TA_t is the current number of tourist arrivals and TA_{t-i} denotes the past realization of total tourist arrivals (i.e. the lagged forecast variable). The optimal lag orders for the forecast variable and the explanatory variables out of a maximum initial lag order equal to 12 are obtained by the BIC through automatic model selection, that is, the ADL models that are taken to the data are obtained through a general-to specific modelling process. The optimal lag orders of the reduced ADL models as determined by the AIC.

3.7 Forecasting Measures

To evaluate the performance of the various models the Root Mean Square Error (RMSE) , the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) are used, which are as follows:

1) $RMSE=\sqrt{MSE}$ where $MSE=\frac{1}{N}\sum_{t=1}^{N}\left(y_t-f_t\right)^2$

2) $MAE=\frac{1}{N}\sum_{t=1}^{N}\left(\left|y_t-x_t\right|\right)$

3) $MAPE=\frac{1}{N}\sum_{t=1}^{N}\left|\frac{\left(y_t-x_t\right)}{y_t}\right|$

4. Empirical results and discussion

4.1 Data

To begin with univariate time series analysis, we focus on tourist arrivals of non-residents in monthly frequency. The data were retrieved from the data site of Cyprus government for the period of January 1990 to December 2019 and are presented in figure 1. Data for the period January 1990 to December 2016 is used as the training sample, and data from January 2017 to December 2019 is used for testing forecasting accuracy. In table 1 we can see the descriptive statistics of our sample.

Figure 1

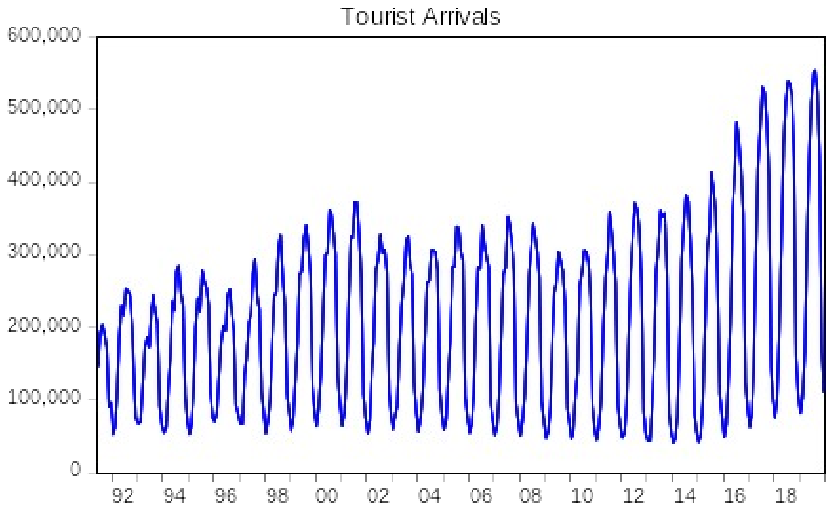


Table 1

Series	Arrivals
Sample	1990M01 2019M12
Observations	360
Mean	203383.7
Median	192045.0
Maximum	553845.0
Minimum	16748.00
Std. Dev.	123594.8
Skewness	0.585151
Kurtosis	2664911
Jarque-Bera	2222838

Key role in our analysis play the terms of stationarity and seasonality. Figure 1 indicates the strong presence of seasonality in tourist arrivals in Cyprus. In addition, the graph shows that there is an upward trend in tourist arrivals from 1990 to 2019. In this stage we test the stationarity of tourist arrivals time series. In this purpose we use the method of Augmented-Dickey Fuller. The results in table 2 demonstrates that the tourist arrivals time series is not stationary , since t-statistic is lower than its critical value and pvalue is higher than 10%. We take our series in natural logarithms. The figure 2 indicates that the seasonality exists. The upward trend also exists but it is not so intense. Applying ADF test in logarithms of tourist arrivals, we observe that the time series is stationary at 5% level of significance. If we take the first differences of natural logarithms of tourist arrivals, we see that there is stationarity with even at 1% level of significance (p-value =0).The first logarithmic differences are very useful in economics because this technique gives the growth rates of interest variables. In addition, taking first logarithmic differences we can remove the trend from our time series.

Table 2. Unit root test

	Tourist Arrivals	
	t-Statistic	P-value
ADF test	-1.355313	0.8721

Figure 2. Tourist Arrivals in Logarithmic form

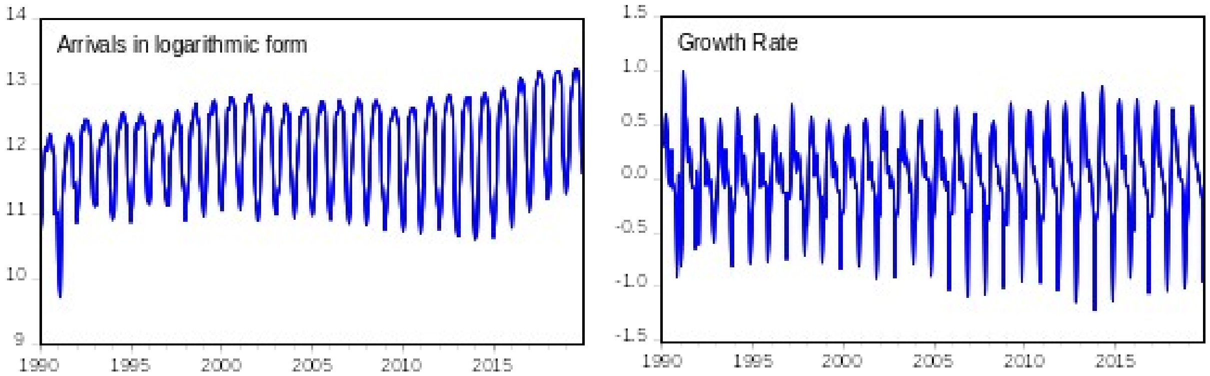


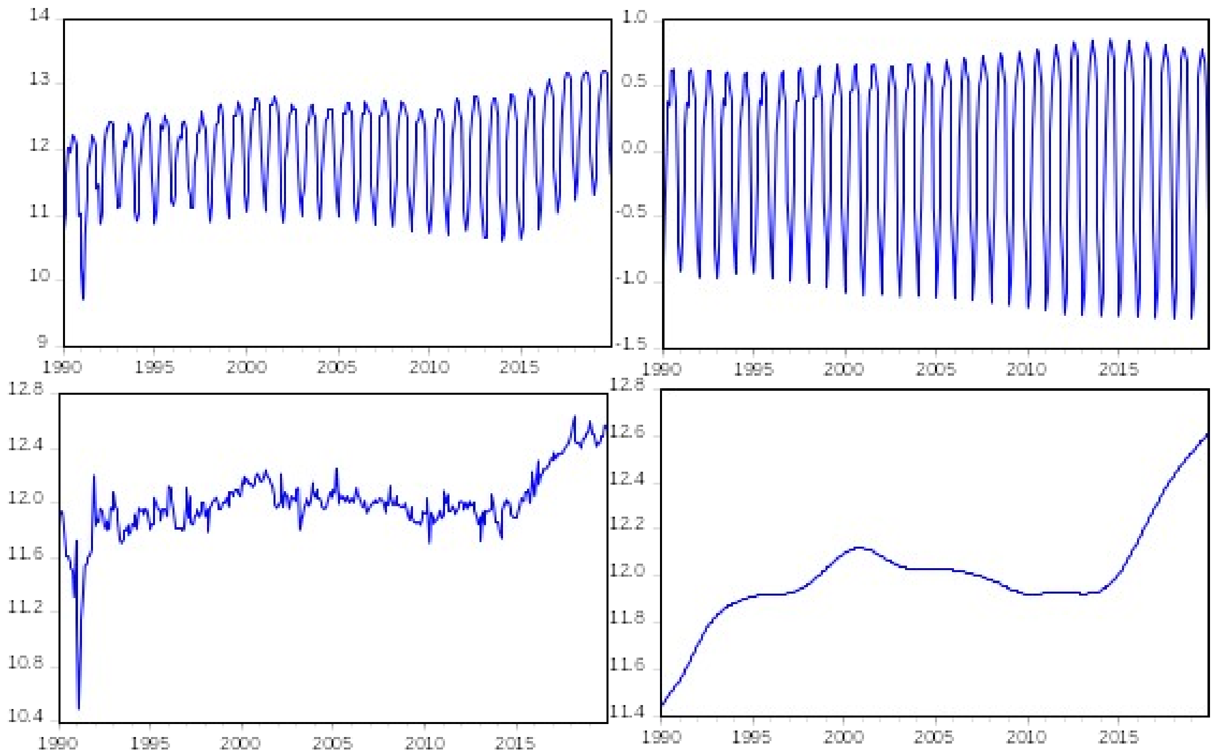
Table 3. Unit root test

	Tourist Arrivals in Logarithmic Form			
	Level		1st Difference	
	t-Statistic	P-value	t-Statistic	P-value
ADF test	-1.355313	0.8721	-1.355313	0.8721

In our investigation we use a variety of econometric methods in order to make forecasts about tourist arrivals in Cyprus and to choose the most efficient model for tourist demand predictions. Desiring to shrink the scale of our data we select to work in logarithmic time series for tourist arrivals. It is apparent that the tourism in Cyprus presents strong evidence of seasonality. It seems that in summer months the tourist arrivals increase in comparison with winter months. Some of the methods that we use do not take into consideration the existence of seasonality. As a result if we ignore this characteristic of our time series, then we will possibly make bias forecasts. For this reason we remove the seasonality through

CENSUS X13 seasonal adjustment method. In the next graphs we see the graphs of raw data, deseasonalized data, seasonality and trend.

Figure 3 Logarithmic form of Arrivals, Seasonality, Deseasonalized series and Trend.



4.2 Results

4.2.1 Simple Models

Table 3. Forecast Evaluation 1

Model	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
Naive	104598.8	77380.64	36.78503	31.73708	0.141697	1.000000
AR(1)	15634.57	10093.95	3.635489	3.611234	0.021156	0.173501
Moving Average (3)	**13817.78**	**8292.031**	**2.760434**	**2.722949**	**0.018776**	**0.145081**
Moving Average (6)	17142.19	10127.60	3.284813	3.227839	0.023302	0.171541
Moving Average (12)	15429.53	8941.126	3.105921	3.034442	0.020799	0.166428
Random Walk	15606.14	10209.77	3.679601	3.661741	0.021136	0.172684
Linear Regression	27131.91	19928.09	6.811376	6.844428	0.036712	0.247985

In the table above we can observe that all forecast accuracy measures indicate the model which makes the most accurate forecasts is the Moving Average model with three terms (MA(3)). In addition, it seems that the Moving Average model with twelve terms, has a better forecast performance in comparison with the naïve, the AR(1) and Random Walk models.

Moreover, the forecasts from linear regression of tourist arrivals on seasonal dummies and trend present the worst predictions compared with the other six models.

4.2.2 Exponential Smoothing

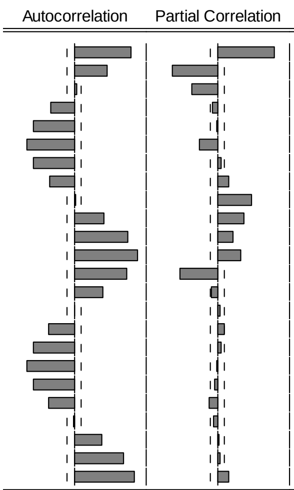
Table 4. Forecast Evaluation 2

Model	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
Single	15766.27	9357.005	3.362986	3.303050	0.021260	0.179136
Double	19071.09	11748.07	3.869138	3.788202	0.025897	0.206886
Additive	**11484.77**	**7783.945**	**2.905486**	**2.902852**	**0.015547**	**0.139073**
Multiplicative	12123.66	8017.674	3.005632	2.999840	0.016410	0.147001

The results demonstrate that the Holt-Winter methods make more accurate forecasts in comparison with the single and the double exponential smoothing methods. The single exponential smoothing is applied in time series without trend and seasonality while the double exponential smoothing in time series without seasonality. Consequently, before we forecast the tourist arrivals in Cyprus through these two methods, we remove the trend and the seasonality from our time series in the first case and only the trend in the second case. On the other hand the Holt-Winter methods take into consideration the trend and the seasonality through their triple exponential smoothing framework and therefore we apply these forecast methods in the raw data (Tourist arrivals in logarithmic form). In table 5 we see that the Holt-Winters exponential smoothing method with additive seasonal variation make the most accurate forecast compared to other three exponential smoothing methods.

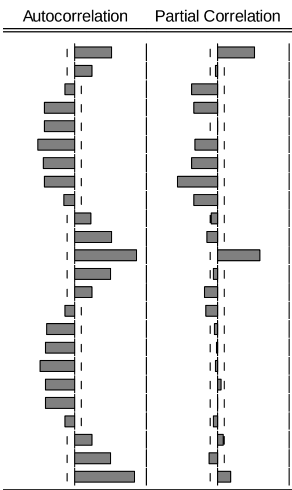
4.3 Auto Regressive Integrated Moving Average (ARIMA- SARIMA)

Figure 4. Correlogram at Level



	AC	PAC	Q-Stat	Prob
1	0.826	0.826	247.89	0.000
2	0.476	-0.653	330.27	0.000
3	0.042	-0.377	330.90	0.000
4	-0.343	-0.080	374.02	0.000
5	-0.583	-0.002	498.98	0.000
6	-0.677	-0.270	667.67	0.000
7	-0.593	0.055	797.66	0.000
8	-0.360	0.170	845.70	0.000
9	0.018	0.498	845.83	0.000
10	0.442	0.382	918.45	0.000
11	0.774	0.224	1142.0	0.000
12	0.924	0.331	1462.0	0.000
13	0.764	-0.543	1681.0	0.000
14	0.421	-0.092	1747.8	0.000
15	0.002	0.041	1747.8	0.000
16	-0.366	0.099	1798.6	0.000
17	-0.594	0.059	1932.9	0.000
18	-0.681	-0.002	2109.7	0.000
19	-0.598	-0.039	2246.3	0.000
20	-0.371	-0.113	2299.2	0.000
21	-0.006	-0.060	2299.2	0.000
22	0.404	0.024	2362.1	0.000
23	0.720	0.035	2562.8	0.000
24	0.862	0.171	2850.8	0.000

Correlogram at 1st Differences



	AC	PAC	Q-Stat	Prob
1	0.534	0.534	103.08	0.000
2	0.266	-0.026	128.80	0.000
3	-0.136	-0.375	135.58	0.000
4	-0.428	-0.329	202.53	0.000
5	-0.437	0.004	272.28	0.000
6	-0.520	-0.323	371.39	0.000
7	-0.445	-0.373	444.41	0.000
8	-0.435	-0.582	514.29	0.000
9	-0.141	-0.338	521.68	0.000
10	0.248	-0.088	544.59	0.000
11	0.536	-0.141	651.74	0.000
12	0.908	0.622	959.87	0.000
13	0.530	-0.065	1065.1	0.000
14	0.256	-0.186	1089.6	0.000
15	-0.139	-0.163	1097.0	0.000
16	-0.404	-0.044	1158.6	0.000
17	-0.417	-0.005	1224.4	0.000
18	-0.499	-0.018	1319.0	0.000
19	-0.423	0.051	1387.2	0.000
20	-0.417	0.005	1453.6	0.000
21	-0.136	-0.059	1460.7	0.000
22	0.259	0.088	1486.5	0.000
23	0.519	-0.119	1590.2	0.000
24	0.871	0.199	1884.0	0.000

Having tested about stationarity in tourist arrivals, we proceed to observe the characteristics of their Autocorrelation function. The correlogram at levels, demonstrates that there are seasonal patterns in the variable of tourist arrivals. In addition, we observe the strong presence of autocorrelation even at 24 months before. On the other hand, in the autocorrelation function of tourist arrivals in first logarithmic differences, we can also observe that the strong evidence of autocorrelation remains. Moreover, if we observe the Autocorrelation function and the Partial Autocorrelation function, we can see that follow similar pattern. Additionally at 1st, 3rd ,4th, 6th, 7th ,8th ,9th ,11th ,12th and 24th lags in the past both AC and PAC are statistically significant. This evidence leads us to estimate ARIMA type models which include both AR and MA terms. Apart from ARIMA model we proceed to the estimation of AR and MA models distinctively, but we also estimate a SARIMA model that incorporates the seasonality.

Table 5 Forecast Evaluation 3

Model	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
MA(10)	16466.26	13195.61	5.023095	5.146041	0.022498	0.175411
AR(10)	**14152.54**	**9497.639**	**3.592585**	**3.608327**	**0.019133**	**0.161203**
SARIMA(4,1,5)(12,1,12)	17194.80	12653.32	4.686914	4.717771	0.023205	0.187649
ARIMA(5,1,4)	20988.47	16262.92	5.633408	5.835520	0.028953	0.198664

Aided by the Eviews-10, we run all possible ARIMA and SARIMA models. The software has an option in add-ins which is named Automatic Arima selection. Through this procedure we take the AR, MA, ARIMA and SARIMA models that minimized the Akaike information criterion. As we can observe the AR and MA models which minimize the information criterion have 10 AR and 10 MA terms respectively. In addition, in table 6 is illustrated that the ARIMA model that minimizes the information criterion includes 5 AR terms and 4 MA terms, while we have taken the first logarithmic differences of tourist arrivals. On the other hand, the selected SARIMA model incorporates 4 AR terms and 5 MA terms, but also 12 seasonal AR and 12 seasonal MA terms as well. As in the ARIMA model we took the growth rates of tourist arrivals. Knowing the presence of seasonality in our data, we estimate the ARIMA model in deseasonalized series while the 16 SARIMA model is applied in the initial series (with seasonality) because the introduction of seasonal terms captures the problem of seasonality. The AR(10) model seems to have the best forecast performance compared with other three models. The evaluation criteria of our forecasts demonstrates that the SARIMA model produce more accurate forecasts than ARIMA.

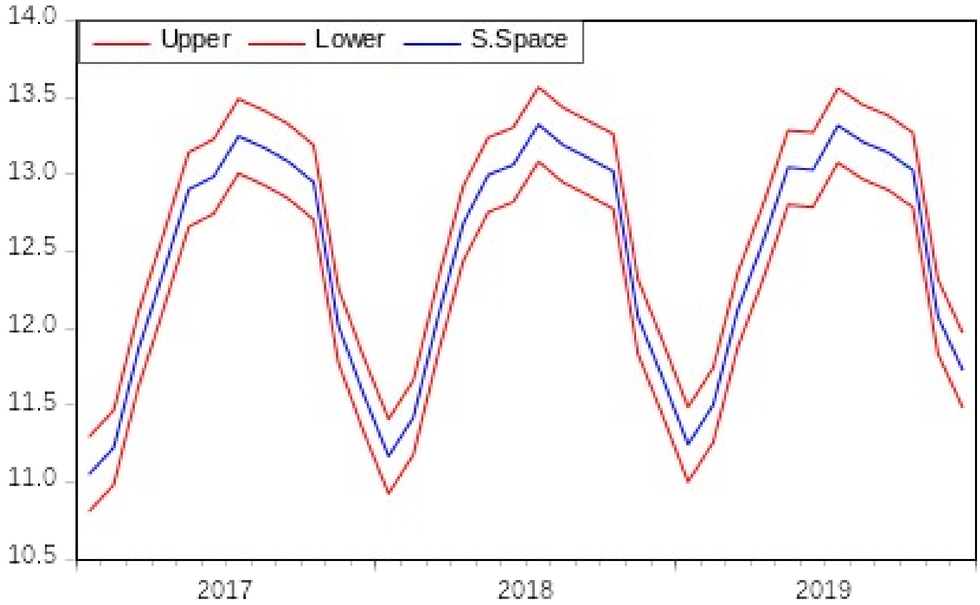
4.4 BSM and SETAR

Table 6 SETAR and Space State Model

Model	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
SETAR	**18298.40**	**15076.17**	**5.321623**	**5.496184**	**0.025152**	**0.179108**
State Space	26421.81	19505.54	6.661671	6.698843	0.035764	0.244892

The results in table 6 demonstrates that the SETAR model produce more accurate forecasts than Unobserved Component Model with the state space formulation. In the graph below we can see the prediction values that we took from the forecast with the State Space Model and their confidence intervals (upper and lower bound).

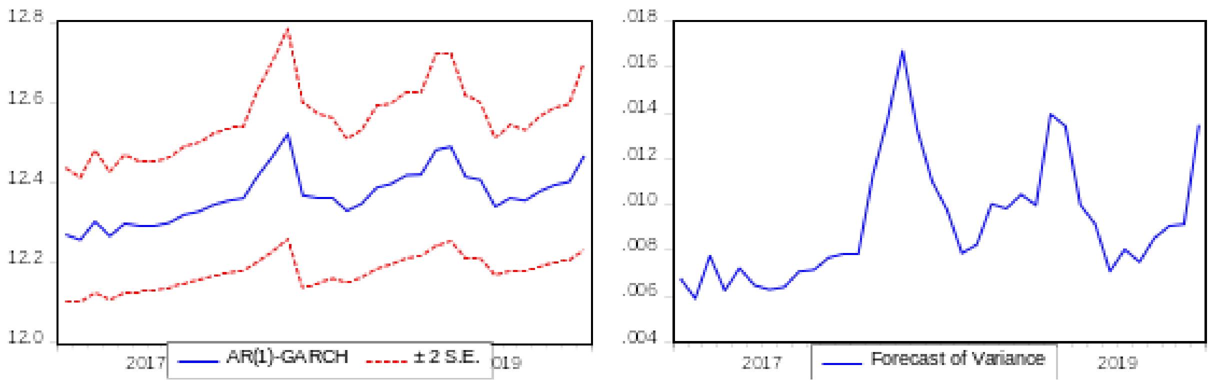
Figure 5. State space model forecast values and confidence intervals



4.5 AR(1)-GARCH AND AR(1) -EGARCH

We use AR(1)-GARCH model in order to forecast tourist arrivals time series. Firstly, the deseasonalized series is used to predict the tourist demand without the presence of seasonality and subsequently we incorporate the seasonal component in our forecast values. In the graphs below we can see not only the forecast values of deseasonalized tourist arrivals time series but also the forecast of tourist demand volatility.

Figure 6. AR(1)-GARCH(1,1)



We follow the same methodology in the prediction of tourist arrivals with a AR(1)- EGARCH model. The predicted values of tourist arrivals deseasonalized series and their volatility forecasts are illustrated in the next graphs.

Figure 7. AR(1)-EGARCH(1,1)

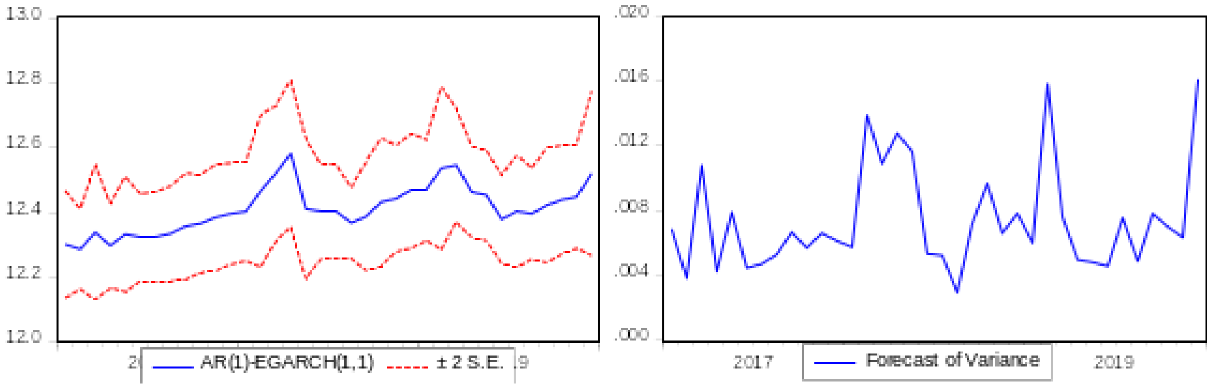


Table 7. Forecasting Evaluation 5

Model	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
AR(1)-GARCH	31224.54	26469.65	8.757386	9.214841	0.043984	0.248641
AR(1)-EGARCH	**20169.92**	**16635.68**	**5.679076**	**5.836853**	**0.027864**	**0.182739**

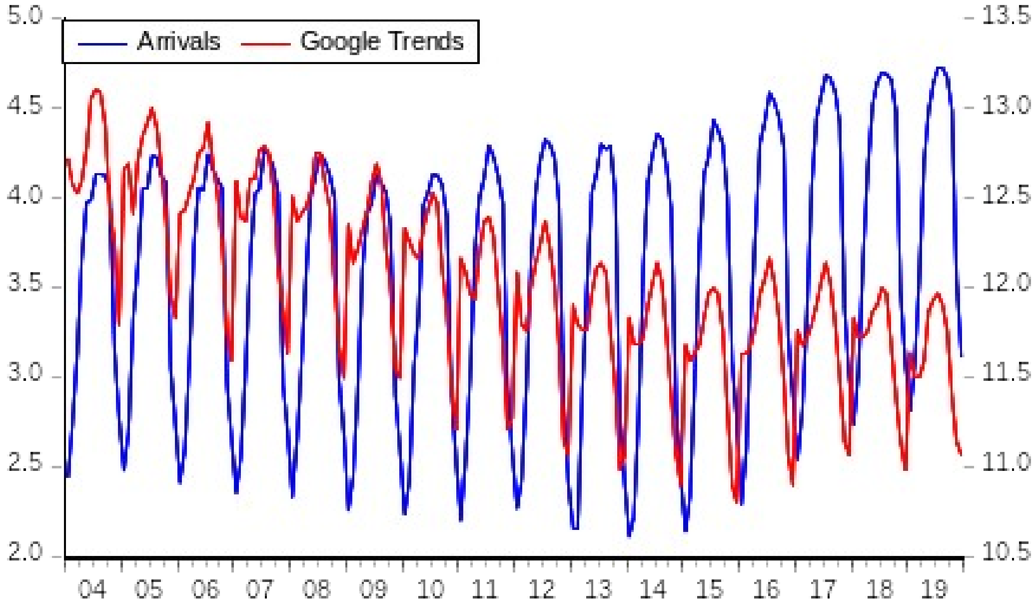
All forecast measures indicate that the asymmetric Garch model with one Autoregressive lag make more accurate forecasts of Cyprus tourist demand in comparison with symmetric AR(1)-GARCH(1,1).

4.6 Tourist Arrivals Prediction through the help of Google Searches Engine

4.6.1 Data

The data for Google trends were available in monthly frequency from January 2004. The data for tourist arrivals are in monthly frequency and cover the period from 1990M01 to 2019M12. Consequently, in this part of our paper we use the data for tourist arrivals and Google trend for the period 2004M01- 2019M12. The dataset of tourist arrivals were retrieved from the data site of Cyprus government for the period of January 2004 to December 2019 and are presented in figure 8. The data for Google trends are retrieved from Google trend database ((https://trends.google.com/trends/) for the same period.

Figure 8. Logarithmic Form of Google Trends Index and Tourist Arrivals



We use the Census-X13 Seasonal Adjustment program and we apply the ADF unit root tests for the deseasonalized series.

Table 8. Unit Root Test

	Variable	Level		1st Difference	
		t-Statistic	P-value	t-Statistic	P-value
ADF test	Tourist Arrivals	-0.740050	0.9679	-4.679438	0.0010
	Google Trends	-0.823786	0.9606	-3.767602	0.0039

It is obvious that both tourist arrivals and Google Trends are not stationary in their levels while they are stationary in first logarithmic differences. On the first logarithmic differences of deseasonalized series are used in the estimations with OLS and ADL methods. On the other hand, the first logarithmic differences of seasonal series are used in SARIMAX because this model captures the seasonality with the incorporation of seasonal AR and MA terms.

Table 9. Forecast Evaluation 6

Model	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
ADL	0.049291	0.036464	0.295338	0.295458	0.001976	0.119168
OLS	**0.044863**	**0.028572**	**0.233236**	**0.233078**	**0.001798**	**0.108613**
SARIMAX(2,1,5)(111)	0.236656	0.179771	1.464562	1.461527	0.009500	0.540038

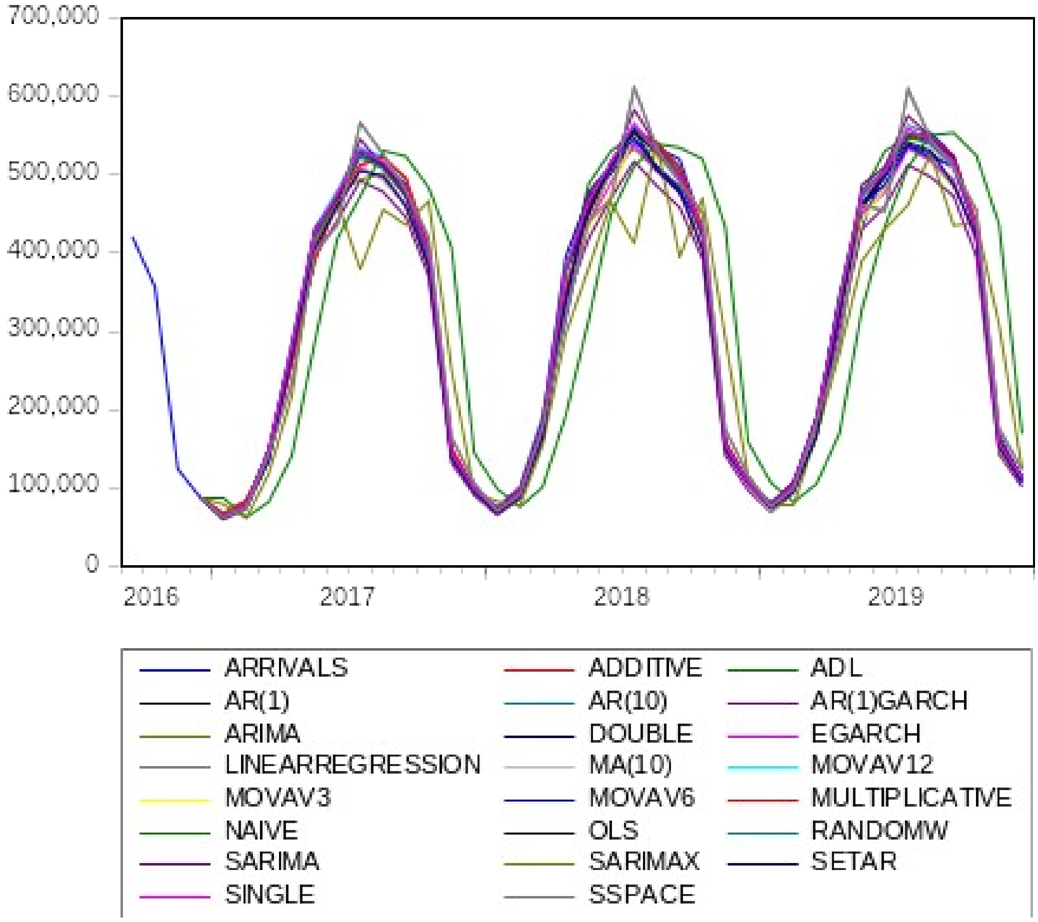
The Seasonal ARIMA model that minimizes the Akaike criterion, includes 2 AR ,5 MA terms 1 seasonal AR and 1 seasonal MA terms. Furthermore, according to Akaike information criterion, the selected ADL model includes 2 lags of dependent variable (tourist arrivals) and 1 lag of independent variable (Google Trends Index). Moreover, the information criterion indicates that we must work in first logarithmic differences (Integrated=1). The outcomes above demonstrate that the forecast from OLS regression is the most accurate. In addition, the forecast evaluation measures indicate that the SARIMAX model shows the worst forecasting performance.

In the next graph we plot the forecasting results from all different methods which we used in the paper. According to RMSE forecasting evaluation measurement, the best and the most appropriate model for forecasting tourist arrivals in Cyprus is the Holt-Winters Exponential Smoothing model with the Additive seasonal effect. The MAPE indicates the Moving Average of three observations as the model with the best forecast accuracy of Tourism demand in Cyprus. Moreover, the OLS regression of tourist arrivals on Google Trends Index seems to have the best forecasting performance according to MAE criterion.

Table 10. Forecast Evaluation 7

Model	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
ADDITIVE	**11484.77**	7783.945	2.905486	2.902852	**0.015547**	**0.139073**
ADL	15350.27	10775.10	3.640689	3.644600	0.020733	0.165634
AR(1)	15634.57	10093.95	3.635489	3.611234	0.021156	0.173501
AR(10)	14152.54	9497.639	3.592585	3.608327	0.019133	0.161203
AR(1)-GARCH	31224.54	26469.65	8.757386	9.214841	0.043984	0.248641
ARIMA54	20988.47	16262.92	5.633408	5.835520	0.028953	0.198664
DOUBLE	19071.09	11748.07	3.869138	3.788202	0.025897	0.206886
LINEARREGRESSION	27131.91	19928.09	6.811376	6.844428	0.036712	0.247985
E-GARCH	20169.92	16635.68	5.679076	5.836853	0.027864	0.182739
MA(10)	16466.26	13195.61	5.023095	5.146041	0.022498	0.175411
MOVAV12	15429.53	8941.126	3.105921	3.034442	0.020799	0.166428
MOVAV3	13817.78	8292.031	**2.760434**	**2.722949**	0.018776	0.145081
MOVAV6	17142.19	10127.60	3.284813	3.227839	0.023302	0.171541
NAIVE	104598.8	77380.64	36.78503	31.73708	0.141697	1.000000
MULTIPLICATIVE	12123.66	8017.674	3.005632	2.999840	0.016410	0.147001
OLS	13563.12	**7735.113**	2.890670	2.855245	0.018374	0.165464
RANDOMW	15606.14	10209.77	3.679601	3.661741	0.021136	0.172684
SARIMA	17194.80	12653.32	4.686914	4.717771	0.023205	0.187649
SARIMAX	66538.07	49628.52	18.78244	17.77912	0.093871	0.489819
SETAR	18298.40	15076.17	5.321623	5.496184	0.025152	0.179108
SINGLE	15766.27	9357.005	3.362986	3.303050	0.021260	0.179136
SSPACE	26421.81	19505.54	6.661671	6.698843	0.035764	0.244892

Figure 9 Forecast Comparison



5.Conclusion

To sum up, in this paper we attempted to find the most accurate econometric model in Cyprus tourist arrivals forecasts and to predict the possible number of arrivals for the next two years. Undeniably the tourism branch constitutes one of the most significant sectors of Cypriot economy and the ability of economic policy makers to estimate the tourist arrivals and to approach the revenues is surely a comparative advantage. The purpose of this paper is to help the economic policy makers and the entrepreneurs of tourism sector to organize their economic and marketing strategies and in this way to promote the further development of the branch in the island. In the first part of our investigation, we gave some general information about the Tourism branch and its contribution in Cypriot economy. In section 2, we presented the econometric methodology which was used in part 3. Key role in our analysis played the seasonality which is obviously observed in Cyprus tourism demand. Using in total 22 different econometric models and the common forecasting accuracy measurements (RMSE, MAE, MAPE) we came to the conclusion that the most suitable models in our case are the Holt-Winter Exponential Smoothing with the additive seasonal effect and the Moving Average model with three terms according to RMSE and MAE respectively.

As George Box (1979) famously and correctly noted, “All models are false, but some are useful.” Precisely the same is true of assumptions, as models are just sets of assumptions.

REFERENCES

Ali, Y., Ciaschini, M., Pretaroli, R., Severini, F., & Socci, C. (2014). Economic relevance of Tourism industry: the Italian case. Quaderno di Dipartimento, (72).

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. International Journal of Forecasting, 27(3), 822-844.

Archer, B. (1987). Demand forecasting and estimation. Demand forecasting and estimation., 77-85.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Bull, A. (1995). Economics of travel and tourism (No. Ed. 2). Longman Australia Pty Ltd.

Cho, V. (2001). Tourism forecasting and its relationship with leading economic indicators. Journal of Hospitality & Tourism Research, 25(4), 399-420.

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

Frechtling, D. C. (1996). Practical tourism forecasting. Butterworth-Heinemann Ltd

Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. Tourism management, 23(5), 499-510.

Gunter, U., Önder, I., & Gindl, S. (2019). Exploring the predictive ability of LIKES of posts on the Facebook pages of four major city DMOs in Austria. Tourism Economics, 25(3), 375-401.

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.

Lundberg, D. E., Krishnamoorthy, M., & Stavenga, M. H. (1995). Tourism economics. John Wiley and sons.

Morley, C. (1991). Modeling international tourism demand: Model specification and structure. Journal of Travel Research, 30(1), 40-44.

Saayman, A., & Botha, I. (2017). Non-linear models for tourism demand forecasting. Tourism Economics, 23(3), 594-613.

Smith, S. L. (2014). Tourism analysis: A handbook. Routledge.

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

Suna, S., Bi, D., Guo, J. E., & Wang, S. (2020). Seasonal and Trend Forecasting of Tourist Arrivals: An Adaptive Multiscale Ensemble Learning Approach.

Xiong, Z., Liu, J., Sriboonchitta, S., & Ramos, V. (2018, July). Forecasting China's inbound tourist arrivals using a state space model. In Journal of Physics: Conference Series (Vol. 1053, No. 1, p. 012134). IOP Publishing.

Yilmaz, E. (2015). Forecasting tourist arrivals to Turkey. Turizam: međunarodni znanstveno-stručni časopis, 63(4), 435-445.