Forecasting Accuracy Evaluation of Tourist Arrivals: Evidence from Parametric and Non-Parametric Techniques

Hossein Hassani* Emmanuel Sirimal Silva† Nikolaos Antonakakis‡§¶
George Filis‖ Rangan Gupta**

July 16, 2015

Abstract

This paper evaluates the use of several parametric and nonparametric forecasting techniques for predicting tourism demand in selected European countries. ARIMA, Exponential Smoothing (ETS), Neural Networks (NN), Trigonometric Box-Cox ARMA Trend Seasonal (TBATS), Fractionalized ARIMA (ARFIMA) and both Singular Spectrum Analysis algorithms, i.e. recurrent SSA (SSA-R) and vector SSA (SSA-V), are adopted to forecast tourist arrivals in Germany, Greece, Spain, Cyprus, Netherlands, Austria, Portugal, Sweden and United Kingdom. This paper not only marks the introductory application of the TBATS model for tourism demand forecasting, but also marks the first instance in which the SSA-R model is effectively utilized for forecasting tourist arrivals. The data is tested rigorously for normality, seasonal unit roots and break points whilst the out-of-sample forecasts are tested for statistical significance. Our findings show that no single model can provide the best forecasts for any of the countries considered here in the short-, medium- and long-run. Moreover, forecasts from NN and ARFIMA models provide the least accurate predictions for European tourist arrivals, yet interestingly ARFIMA forecasts are better than the powerful NN model. SSA-R, SSA-V, ARIMA and TBATS are found to be viable options for modelling European tourist arrivals based on the most number of times a given model outperforms the competing models in the above order. The results enable forecasters to choose the most suitable model (from those evaluated here) based on the country and horizon for forecasting tourism demand. Should a single model be of interest, then, across all selected countries and horizons the SSA-R model is found to be the most efficient based on lowest overall forecasting error.

Keywords: Tourist arrivals; Tourism demand; Forecasting; Singular Spectrum Analysis; ARIMA; Exponential Smoothing; Neural Networks; TBATS; ARFIMA.

*Statistical Research Centre, Bournemouth University, 89 Holdenhurst Road, Bournemouth BH8 8EB, UK, email: hhassani@bournemouth.ac.uk.

†Statistical Research Centre, Bournemouth University, 89 Holdenhurst Road, Bournemouth BH8 8EB, UK, email: esilva@bournemouth.ac.uk.

‡Vienna University of Economics and Business, Department of Economics, Institute for International Economics, Welthandelsplatz 1, 1020, Vienna, Austria, e-mail: nikolaos.antonakakis@wu.ac.at.

§University of Portsmouth, Economics and Finance Subject Group, Portsmouth Business School, Portland Street, Portsmouth, PO1 3DE, United Kingdom, e-mail: nikolaos.antonakakis@port.ac.uk.

¶Johannes Kepler University, Department of Economics, Altenbergerstrae 69, Linz, 4040, Austria, e-mail: nikolaos.antonakakis@jku.at.

‖Bournemouth University, Accounting, Finance and Economics Department, 89 Holdenhurst Road, Bournemouth, Dorset, BH8 8EB, United Kingdom, e-mail: gfilis@bournemouth.ac.uk.

**Department of Economics, Faculty of Economic and Management Sciences, University of Pretoria, 0002, South Africa, email: rangan.gupta@up.ac.za.

1 Introduction

Tourism in the 21th century has experienced continued expansion and diversification, becoming one of the largest and fastest-growing economic sectors in the world. Among the most favourite destinations, Europe is considered the most prominent one, receiving the highest amount of tourists arrivals (563 million), representing 52% of the global tourist arrivals and generating an income of more than €368 billions in 2013 (UNWTO, 2014). However, despite Europe being the region with the most arrivals, it is not the region that is growing at the fastest rate. According to UNWTO (2014), regions such as Asia and the Pacific, and Africa that have traditionally had a lower rate of arrivals are experiencing the highest growth in recent years. These developments might be due to the global financial crisis and the ongoing European debt crisis that Europe has suffered the most from (e.g., see Antonakakis et al., 2015a,b). Since the European Union has placed a lot of emphasis on the tourism sector as a source of economic prosperity for its member countries (Lee and Brahmasrene, 2013), the need of accurate forecasts of tourism demand is of paramount importance for tourism planning, entrepreneurs, investors, policy makers, tour operators and others alike.

In addition, various time horizons are relevant to decision making in the tourism sector. For example, short-term forecasts are required for scheduling and staffing, while medium-term forecasts for planning tour operator brochures and long-term forecasts for investment in aircraft, hotels and infrastructure.

To that end, the purpose of this study is to evaluate both the short-, medium- and long-run forecasting accuracy of tourism demand based on several parametric and nonparametric forecasting techniques in selected European countries, namely, Austria, Cyprus, Germany, Greece, Netherlands, Portugal, Spain, Sweden and the United Kingdom. In contrast to previous studies, that compare different classes of the same model or a few different classes of models, this study employs seven alternative parametric and non-parametric techniques, thereby complementing all previous studies in an attempt to uncover the best forecasting method of tourist arrivals in Europe. In particular, the models employed include the Autoregressive Moving Average (ARIMA), Exponential Smoothing (ETS), Neural Networks (NN), Trigonometric Box-Cox ARMA Trend Seasonal (TBATS), Fractionalized ARIMA (ARFIMA) and both Singular Spectrum Analysis algorithms, i.e. recurrent SSA (SSA-R) and vector SSA (SSA-V). This study not only marks the introductory application of the TBATS model for tourism demand forecasting, but also marks the first instance in which the SSA-R model is effectively utilized for forecasting tourist arrivals. The TBATS introduced by De Livera et al. (2011) relies on a new method that greatly reduces the computational burden in the maximum likelihood estimation when forecasting complex seasonal time series such as those with multiple seasonal periods, high-frequency seasonality, non-integer seasonality, and dual-calendar effects, while the SSA-R approach of Hassani et al. (2013) is a nonparametric approach that has very good properties in dealing with both stationary as well as non-stationary data and requires no prior assumptions about the data-generating process. The TBATS has been used to forecast energy consumption (Silva and Rajapaksa, 2014), the price of gold (Hassani et al., 2015) and housing downturns (Zietz and Traian, 2014), while the SSA-R to forecast inflation dynamics (Hassani et al., 2013) among others.

Put differently, this study provides the most comprehensive forecasting comparison among several parametric and non-parametric techniques of international tourist arrivals in Europe. Following the comprehensive univariate modelling exercise we go a step further and seek to ascertain leading cross country indicators for tourism demand in the selected European nations. For this purpose we also introduce an automated multivariate SSA based algorithm in this paper. Note that, in this paper, as discussed, we follow an univariate approach to forecasting tourist arrivals. There are two reasons for this: First, as indicated by Antonakakis et al. (2015a,b), on average tourism is a leading indicator for the economies under consideration. In light of this, it is only rational that we try and develop univariate forecasting models for tourist arrivals, which allows us to forecast the same independent of other macroeconomic variables that possibly affects tourist arrivals. Second, the tourism-growth literature (see, for example Arslanturk et al., 2011; Balcilar et al., 2014, and references cited therein for detailed literature reviews) indicates that there are possibly large number of variables that can affect both tourism and growth simultaneously. Given this, at this stage, we avoided possible selection bias in choosing such variables for these countries. However, we leave this as a possible venue of future research, which we discuss further in the conclusion.

Our findings reveal that no single model can provide the best forecasts for any of the countries considered here in the short-, medium- and long-run. Moreover, forecasts from NN, ETS and ARFIMA models provide the least accurate predictions for European tourist arrivals, yet interestingly ARFIMA forecasts are better than the powerful NN model. SSA-R, SSA-V, ARIMA and TBATS are found to be viable options for modelling European tourist arrivals based on the most number of times a given model outperforms the competing models in the above order. The paper also computes information on the ability of the forecasts to predict the correct direction of change in the data which adds value to the overall results. Thus, the nature in which the results have been presented enables forecasters to choose the most suitable model (from those evaluated here) based on the country, horizon and direction of change criteria for forecasting tourism demand. Should a single model be of interest, then, across all selected countries and horizons the SSA-R model is found to be the most efficient based on lowest overall forecasting error.

The remainder of the paper is organised as follows. Section 2 reviews the most related studies on forecasting methods of tourist arrivals. Section 3 discusses the various parametric and non-parametric forecasting techniques employed in this study. Section 4 presents the data used and the measures employed for evaluating forecast accuracy. Section 5 presents the empirical results. Finally, Section 6 concludes this study.

2 Literature Review

Along with the phenomenal growth in demand for tourism in the world over the past two decades, there is a growing interest in tourism forecasting research. The empirical literature on forecasting tourism demand shows that there is not a single model that has superior predictive ability. Rather, a number of different parametric and non-parametric time-series models, as well as, various econometrics models have been applied in this crowded strand of the tourism literature. Although no consensus has been reached so far, regarding the model with the best forecasting accuracy, the literature reveals that the ARIMA-type models are the most widely used ones.

Starting from these models, one of the early studies is this by Dharmaratne (1995) who compares a number of ARIMA-type models to forecast tourist arrivals in Barbados. The study concludes that ARIMA-type models are capable of producing valid forecasts but specifically the ARIMA(2,1,1) is the best performing model.

Furthermore, a number of authors compare ARIMA-type models with other time-series or econometric models (see, inter alia Goh and Law, 2002; Kulendran and Witt, 2003; Chu, 2004; Kim and Moosa, 2005; Vu and Turner, 2006; Wong et al., 2007; Chu, 2008; Brida and Risso, 2011; Wan et al., 2013).

More specifically, Goh and Law (2002) apply Seasonal ARIMA (SARIMA) and Multivariate ARIMA (MARIMA) models, and compare their forecasting accuracy against a number of exponential smoothing models, moving average models, as well as, a random walk model (naive model). The authors argue that both the SARIMA and MARIMA models outperform all remaining models across a number of forecasting accuracy criteria.

Furthermore, Kulendran and Witt (2003) use a number of ARIMA specifications, a causal structural time-series model (STSM), a basic structural model (BSM), as well as, the naive model of no change. The findings suggest that the ARIMA models exhibit superior predictive ability in the short-run forecasts; however, none of the models could outperform the naive model in the medium-run forecasts. Vu and Turner (2006) second the findings by Kulendran and Witt (2003), as they also compare ARIMA models against a BSM for the case of Thailand and find that the ARIMA models showed a better forecasting accuracy.

Chu (2004) further examines whether a cubic polynomial model could outperform other linear and nonlinear forecasting models, such as a regression-base model, two naive models, ARIMA-type models and a sine wave nonlinear model, which have been estimated in the earlier studies of Chan (1993) and Chu (1998a,b). The study focuses on tourist arrivals in Singapore and the findings suggest that the cubic polynomial model cannot outperform either the ARIMA-type models or the combined forecasts.

Another study that confirms the superiority of the ARIMA-type models is this by Kim and Moosa (2005) who compare the SARIMA model against a regression-based model and Harveys structure time-series model. They also compare the forecasting accuracy of these models based on both aggregate and disaggregate data. Their results suggest that the SARIMA models perform better than the other two models. They also claim that disaggregate data offer better predictive ability compared to aggregate data.

Furthermore, Chu (2008) uses nine time-series models, including two naive models, ARIMA-type models (ARIMA, SARIMA and ARFIMA), as well as, regression-based models. Chu (2008) finds that the ARFIMA model that exhibits the highest forecasting accuracy both in the short-run and in the long-run, nevertheless, the SARIMA is the best performing model in the medium-run. In a subsequent paper, Chu (2009) confirms his previous findings, suggesting that the ARFIMA model performs better compared to other ARIMA specifications.

More recently, Wan et al. (2013) use a SARIMA model and compare it against a seasonal moving average model and a Holt-Winter model. Their findings show that the SARIMA model is the best performing under all three different h-step-ahead forecasting horizons (where h is one-month, three-months and twelve-months ahead).

Other authors have tried to combine ARIMA-type models with ARCH-type models. Indicatively, Coshall (2009) combines the ARIMA with the GARCH models and compares their forecasting accuracy against the Holt-Winters additive and multiplicative exponential smoothing, as well as, a naive model. The results show that the Holt-Winters models perform better in the one and three year-ahead forecasts, whereas the ARIMA-GARCH model yields the best forecasts for the two years-ahead horizon. However, forecasts based on the combined models between the ARIMA-GARCH and the Holt-Winters models provide the most accurate forecasts in almost all sample countries and horizons.

Furthermore, Brida and Risso (2011) compare two SARIMA-ARCH models and show that overall SARIMA-ARCH-type models are able to generate accurate forecasts. In particular, the SARIMA(2,1,2)(0,1,1)–ARCH(1) produces the best forecasts.

Despite the fact that a wealth of studies demonstrates the superior predictive ability of the ARIMA-type models, there are studies that cannot subscribe to this belief.For instance, Song et al. (2003) consider six econometric models (including a regression-based model, a Wickens-

Breusch Error Correction Model (ECM), Johansens ECM, an Autoregressive Distributed Lag Model (ADLM), an unrestricted Vector Autoregressive Model (VAR), a Time Varying Parameter model) and two time-series models (ARIMA and naive model of no change) and produce forecasts for one up to four years-ahead. The results show that there is not a single model that outperforms all others across all different forecasting horizons. In particular, the Time Varying Parameter model is the best performing model for the one and two years ahead; nevertheless, for the longer-term forecasts it is the regression-based model that has the best forecasting accuracy.

Similarly, Wong et al. (2007) compare the ARIMA models with several other time-series and econometric models, such as the ADLM, ECM and VAR. The authors cannot confirm the superiority of the ARIMA models or any other model over the others, for all sample countries. What is more, the authors suggest that in some cases the best forecasting accuracy can be obtained with combined forecast models.

Chu (2011) also uses AR, ARFIMA and SARIMA models and compares them against the forecasting power of a piecewise linear model for Macaus tourism demand. Focusing on four different forecasting horizons (spanning from 6 months to 24 months), they claim that the piecewise linear model is able to outperform all other benchmark models for all forecasting horizons.

More recently, Kim et al. (2011) cannot confirm the superior forecasting accuracy of the ARIMA-type models. More specifically they consider SARIMA models and compare them with autoregressive models (AR), Harveys structural time series model, state space exponential smoothing models and a forecasting model with bootstrap bias-corrected AR parameters. They report that the latter model has superior predictive ability.

At the same time, there are studies which did not consider the ARIMA models at all. For instance, Lim and McAleer (2001) use a number of Holt-Winters and Browns exponential smoothing models, such as the single and double exponential smoothing, non-seasonal and seasonal exponential smoothing, as well as, additive and multiplicative Holt-Winters seasonal smoothing.Theyreport that the Holt-Winters multiplicative seasonal model outperforms all other specifications for the majority of the countries under examination. They note that, in some cases, the Holt-Winters additive seasonal model yields the best forecasts.

Along the same lines, Wong et al. (2006) do not consider any ARIMA model but rather concentrate on various VAR models, including both unrestricted and Bayesian (BVAR) models. They conclude that the univariate BVAR model outperforms all other specifications, including the standard and the general BVAR models. Similarly, Song and Witt (2006) focus only on VAR models and maintain that these models are capable of producing valid forecasts at both the medium- and long-run.

Furthermore, there are studies which turn their attention to biological algorithms in an effort to achieve greater forecasting accuracy for tourism demand. One of the first studies is this by Law and Au (1999) who use a supervised feed-forward neural network to forecast tourist arrivals in Japan. Their findings show that the use of the neural network model is able to outperform the forecasts produced by regression-based models, naive models or even those produced by exponential smoothing and moving average models.

Cho (2003) also use neural network models and compares them against ARIMA and exponential smoothing models. This study confirms the superior character of the neural network model, which was the model that showed the best forecasting accuracy. Burger et al. (2001), in an earlier study, also show that aneural network model can outperform the ARIMA models, as well as, various exponential smoothing, regression-based and naive models.

Furthermore, Kon and Turner (2005) compare a neural network model against a basic structural method in order to identify whether the former can outperform the forecasting accuracy of the latter for the tourist arrivals in Singapore. The authors also use two more models as benchmarks, namely, a naive model and a Holt-Winters model. The findings show that a well structured neural network model can outperform all other models for short-run forecasts.

Other recent studies that focus on biological algorithms include these by Palmer et al. (2006), Hadavandi et al. (2011) and Pai et al. (2014). More specifically, Palmer et al. (2006) develop an artificial neural network (ANN) to forecast tourism arrivals and they claim that an ANN can perform better compared to the traditional statistical models. In addition, Hadavandi et al. (2011) apply a genetic fuzzy system (GFS) and show that biological algorithms are capable of producing successful forecasts for tourism arrivals. Furthermore, Pai et al. (2014) use a fuzzy c-means model with least-square support vector regression algorithm. They report that the use of such hybrid system is a promising alternative for tourism arrivals forecasts compared to standard forecasting models, such as ARIMA.

On the contrary, Claveria and Torra (2014) do not agree with these aforementioned findings, showing that neural networks cannot outperform the ARIMA models, especially for the short-run forecasts.

Summing up, the empirical literature has provided mixed results in terms of tourism demand forecasting accuracy among the various employed models that reveal several idiosyncratic features, both in terms of the forecasting horizons and countries of interest.

3 Forecasting Methods

3.1 Auto-Regressive Integrated Moving Average (ARIMA)

This paper exploits an optimized version of the ARIMA model which is found in the forecast package in *R*. Those interested in a detailed description of the algorithm are referred to Hyndman and Khandakar (2008). The number of seasonal differences, *d*, and the the determination of its value is based on the Osborn-Chui-Smith-Birchenhall test (Osborn et al. 1988) seasonal unit root test. Then, the Akaike Information Criterion (AIC) of the following form is minimized to determine the values of p and q.

$$AIC = -2log(L) + 2(p + q + P + Q + k), \tag{1}$$

where $k = 1$ if c $\neq 0$ and 0 otherwise and L is the maximum likelihood of the fitted model.

Then, the algorithm searches for the model which represents the smallest AIC from: ARIMA*(2,d,2)*, ARIMA*(0,d,0)*, ARIMA*(1,d,0)* and ARIMA*(0,d,1)* which is selected as the optimal ARIMA model. The decision on the inclusion or exclusion of the constant c depends on the value of d. As seen in the next section, all time series considered in this study have a seasonal unit root problem and therefore we provide a brief expansion of the seasonal ARIMA model alone. In doing so we mainly follow Hyndman and Khandakar (2008). Accordingly, the seasonal ARIMA model can be expressed as:

$$\Phi(B^m)\phi(B)(1 - B^m)^D(1 - B)^d y_t = c + \Theta(B^m)\theta(B)\epsilon_t, \tag{2}$$

where $\Phi(z)$ and $\Theta(z)$ are the polynomials of orders P and Q, and ϵ_t is white noise. If, $c \neq 0$, there is an implied polynomial of order $d + D$ in the forecast function.

As explained in Hyndman and Athanasopoulos (2013) point forecasts can then be obtained as follows. Begin by expanding the seasonal ARIMA equation so that y_t is on the left hand side with all other terms on the right. Then, rewrite the ARIMA equation and replace t with $T + h$ and finally, on the right hand side of this equation replace future observations by their forecasts, future errors by zero, and past errors by the corresponding residuals. Eventually, use the forecasting horizon $h = 1$ month ahead for example to calculate all forecasts for that horizon.

3.2 Exponential Smoothing (ETS)

In brief, the ETS model considers the error, trend and seasonal components in choosing the best exponential smoothing model from over 30 possible options by optimizing initial values and parameters using the MLE for example and selecting the best model based on the AIC. This ETS algorithm overcomes limitations from the previous models of exponential smoothing which failed to provide a method for easily calculating prediction intervals (Makridakis, Wheelwright, & Hyndman 1998). Those interested in a detailed description of ETS are referred to Hyndman and Athanasopoulos (2013).

3.3 Neural Networks (NN)

The neural network models used in this paper are estimated using an automatic forecasting model known as nnetar which is provided through the forecast package in R programming code. For a detailed explanation on how the nnetar model operated, see Hyndman et al. (2013). The parameters in the neural network model are selected based on a loss function embedded into learning algorithm. The nnetar algorithm trains 25 networks by using random starting values and then obtains the average of the resulting predictions to compute the forecast. It may be noted that in all cases the selected neural network model has only k=1 hidden node, p=2 lags and we adopt annual difference specifications. Thus, for these series it appears that simpler network models perform better than more complex ones.

3.4 Trigonometric Box-Cox ARMA Trend Seasonal Model (TBATS)

The TBATS model is an exponential smoothing state space model with Box-Cox transformation, ARMA error correction, Trend and Seasonal components. The result is a technique which is aimed at providing accurate forecasts for time series with complex seasonality. A detailed description of the TBATS model can be found in De Livera et al. (2011).

3.5 Fractionalized ARIMA Model (ARFIMA)

The ARFIMA modelling process provided through the forecast package in R automatically estimates and selects p and q for an ARFIMA(p,d,q) model based on the Hyndman and Khandakar (2008) algorithm whilst d and parameters are selected based on the Haslett and Raftery (1989) algorithm.

3.6 Singular Spectrum Analysis (SSA)

The basic SSA technique is well established and detailed in literature. Those interested in a detailed description of the two main stages of SSA (i.e. Decomposition and Reconstruction), are directed to Hassani (2007); Golyandina et al. (2001). Figure 1 presents a summary of the basic SSA process. Thereafter the SSA-R and SSA-V forecasting algorithms are concisely explained.

[Insert Figure 1 around here]

SSA-R

Let $v^2 = \pi_1^2 + \ldots + \pi_r^2$, where π_i is the last component of the eigenvector U_i $(i = 1, \ldots, r)$. Moreover, suppose for any vector $U \in \mathbf{R}^L$ denoted by $U^\triangledown \in \mathbf{R}^{L-1}$ the vector consisting of the first $L-1$ components of the vector U. Let $y_{N+1}, \ldots, y_{N+h}$ show the h terms of the SSA recurrent forecast. Then, the h-step ahead forecasting procedure can be obtained by the following formula

$$y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, N \\ \sum_{j=1}^{L-1} \alpha_j y_{i-j} & \text{for } i = N+1, \ldots, N+h \end{cases} \tag{3}$$

where $\widetilde{y}_i$ $(i = 1, \ldots, N)$ creates the reconstructed series (noise reduced series) and vector $A = (\alpha_{L-1}, \ldots, \alpha_1)$ is computed by:

$$A = \frac{1}{1 - v^2} \sum_{i=1}^{r} \pi_i U_i^\triangledown. \tag{4}$$

SSA-V

Consider the following matrix

$$\Pi = \mathbf{V}^\triangledown (\mathbf{V}^\triangledown)^T + (1 - v^2) A A^T \tag{5}$$

where $\mathbf{V}^\triangledown = [U_1^\triangledown, ..., U_r^\triangledown]$. Now consider the linear operator

$$\theta^{(v)} : \mathfrak{L}_r \mapsto \mathbf{R}^L \tag{6}$$

where

$$\theta^{(v)} U = \begin{pmatrix} \Pi U^\triangledown \\ A^T U^\triangledown \end{pmatrix}. \tag{7}$$

Define vector Z_i as follows:

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, K \\ \theta^{(v)} Z_{i-1} & \text{for } i = K+1, \ldots, K+h+L-1 \end{cases} \tag{8}$$

where, $\widetilde{X}_i$'s are the reconstructed columns of the trajectory matrix after grouping and eliminating noise components. Now, by constructing matrix $\mathbf{Z} = [Z_1, ..., Z_{K+h+L-1}]$ and performing diagonal averaging we obtain a new series $y_1, ..., y_{N+h+L-1}$, where $y_{N+1}, ..., y_{N+h}$ form the h terms of the SSA vector forecast.

4 The Data and Measures for Evaluating Forecast Accuracy

4.1 The Data

This papers focuses on international tourist arrivals in European countries, namely, Austria, Cyprus, Germany, Greece, Netherlands, Portugal, Spain, Sweden and the United Kingdom. The data on international tourist arrivals is obtained from Eurostat database. The period spans from January 2000 until December 2013.

We begin our analysis by testing the data for normality, seasonal unit roots and break points. From the descriptive statistics reported in Table 1, the Shapiro-Wilk (SW) test for normality indicates that tourist arrivals in Austria is the only normally distributed series. This suggests that when discussing central tendency and variation it is more appropriate to consider the median and IQR for all majority of the series which are skewed whilst for Austrian tourist arrivals the mean and standard deviation (SD) criterion is appropriate. During the 13 year period, the highest median tourist arrivals was reported in Italy whilst the lowest median tourist arrivals had been in Cyprus. Based on the IQR, the most variation in tourist arrivals was recorded in Italy whilst the least variation was in Cyprus. However, if we were to consider variation in monthly tourist arrivals based on the standard deviation then again the results are consistent with those reported based on the IQR. As majority of the tourist arrivals series are skewed, it is better to rely on the coefficient of variation (CV) criterion to compare the variability between countries. Based on the CV, Greece reports the highest variation in tourist arrivals whilst Netherlands reports the lowest variation in tourist arrivals. The OCSB (Osborn et al. 1988) test for seasonal unit roots indicates that except for the Dutch tourist arrivals series, all other series have seasonal unit roots.

[Insert Table 1 around here]

Results from the Bai and Perron (2003) test for break points is reported in Table 2. Between 2000 and 2013 the only country to experience two structural breaks in tourist arrivals is Germany whilst Cyprus and Sweden has experienced no structural breaks during this period. We use this information to determine training and validation sets for our forecasting exercise which follows. As 2011 April is the last structural break experienced by at-least one of the countries considered here, we use data from January 2000 - April 2011 for training and testing the forecasting models, and set aside as validation sets the observations from May 2011 - December 2013 which is approximately 2.5 years. This is done in order to ensure that no model has any undue advantage because both parametric and nonparametric methods are considered in this study. It is well known that methods such as SSA can handle non-stationarity well, and that it is less sensitive to structural breaks as was shown recently in Silva and Hassani (2015) where the authors considered the same ARIMA, ETS and Neural Network models from the forecast package alongside SSA in an application on forecasting U.S. trade. Moreover, this approach will enable us to ascertain whether structural breaks in the training samples have adverse effects on the forecasts generated by these models.

[Insert Table 2 around here]

Reported in Table 3 are the parameters of the fitted models during the training process for the selected European tourist arrivals series. It should be noted that the parameters reported for ARIMA, ETS, NN, TBATS and ARFIMA are those relevant at the first instance. This is because these parameters keep changing over any selected forecasting horizon as the algorithms re-estimate a new model fit each time a new observation is introduced. In contrast, the SSA-V and SSA-R model parameters once fitted remain constant and do not vary. Thus, the SSA model is more stable and it will be interesting to see how the constant SSA models compete with the varying models in the forecasting exercise which follows.

[Insert Table 3 around here]

Whilst ARIMA, ETS, NN, TBATS, and ARFIMA models can be automatically estimated via the algorithms freely accessible through the forecast package in *R*, for SSA, here we use the conventional method which requires an understanding of the theory underlying the technique. As such, in order to enlighten the reader on how each SSA model was trained, as an example, below we present Figure 2 and briefly explain the process involved in fitting the SSA(48,13) model for German tourist arrivals.

[Insert Figure 2 around here]

We begin by considering only the training data for German tourist arrivals. The first step is to analyze the periodogram in order to identify the dominating frequencies. In this case it is clear that the 12 month seasonal component is dominating tourist arrivals in Germany with some comparatively small peaks visible around 2, 4 and 6 months as well. Accordingly, we follow the method in Hassani (2007) and select L proportional to the dominating frequency of 12. We then evaluate L = 24, 36, 48 and 60, and during each evaluation we study the paired-eigenvectors to ascertain which decomposition provides the best in-sample fitting. The paired-eigenvectors for Germany showed that beyond $r = 13$ there were no eigenvectors which represents the seasonal components of interest, i.e. 12, 6, 4, and 2 months. As such, in this case we choose SSA(48,13) as the fitted model for Germany. This model is then used to calculate the out-of-sample forecasts. The same steps are followed for the remaining time series.

As SSA is the only filtering technique used in this paper we find it pertinent to comment with regard to the separation of signal and noise as achieved via SSA. The weighted correlation (w-correlation) statistic can be used to present the appropriateness of the various decompositions achieved by SSA (see, Table 3). As mentioned in Golyandina, et al. (2001), the w-correlation is a statistic which shows the dependence between two time series. It can be calculated as:

$$\rho_{12}^{(w)} = \frac{\left(Y_N^{(1)}, Y_N^{(2)}\right)_w}{\parallel Y_N^{(1)} \parallel_w \parallel Y_N^{(2)} \parallel_w},$$

where $Y_N^{(1)}$ and $Y_N^{(2)}$ are two time series, $\parallel Y_N^{(i)} \parallel_w = \sqrt{\left(Y_N^{(i)}, Y_N^{(i)}\right)_w}$, $\left(Y_N^{(i)}, Y_N^{(j)}\right)_w = \sum_{k=1}^{N} w_k y_k^{(i)} y_k^{(j)} \quad (i,j = 1,2)$, w_k=min$\{k, L, N-k\}$ (here, assume $L \leq N/2$).

The w-correlation is interpreted such that if its value between two reconstructed components are close to 0, it confirms that the corresponding time series are w-orthogonal and are well separable (Hassani et al. 2009), and thus confirms the noise is indeed random even though residual randomness is not an explicit concern for nonparametric models. Table 4 shows the w-correlations for all SSA decompositions by comparing the two components of signal and noise. Here, we use as signal the reconstructed series containing r components and select the remaining r (which does not belong to the reconstruction) as noise. As evident, all w-correlations are close to 0 and this confirms that SSA has successfully achieved a sound separation between noise and signal.

[Insert Table 4 around here]

4.2 Measures for Evaluating the Forecast Accuracy

Root Mean Squared Error (RMSE)

The RMSE is used to measure the forecast accuracy. Recently it has been widely adopted in forecasting literature, see for example, Zhang et al. (1998), Hassani et al. (2009;2013;2015). Here, in order to save space, we only provide the RMSE ratios of SSA to that of NN:

$$\text{RMSE} = \frac{SSA}{NN} = \frac{\left(\sum_{i=1}^{N}(\widehat{y}_{T+h,i} - y_{T+h,i})^2\right)^{1/2}}{\left(\sum_{i=1}^{N}(\widetilde{y}_{T+h,i} - y_{T+h,i})^2\right)^{1/2}},$$

where, $\widehat{y}_{T+h}$ is the h-step ahead forecast obtained by SSA, $\widetilde{y}_{T+h}$ is the h-step ahead forecast from the NN model, and N is the number of the forecasts. If $\frac{SSA}{NN}$ is less than 1, then the SSA outperforms NN by 1-$\frac{SSA}{ETS}$ percent.

Direction of Change (DC)

The DC criterion is a measure of the percentage of forecasts that accurately predict the direction of change (Hassani, Heravi, & Zhigljavsky, 2013). DC is an equally important measure, as the RMSE, for evaluating the forecasting performance of tourism demand models, because it is important that for example, when the actual series is illustrating an upwards trend, the forecast is able to predict that upward trend and vice versa. Here, the concept of DC is explained in brief, and in doing so we mainly follow Hassani, Heravi, and Zhigljavsky (2013). In the univariate case, for forecasts obtained using X_T, let D_{Xi} be equal to 1 if the forecast is able to correctly predict the actual direction of change and 0 otherwise. Then, $\tilde{D}_X = \sum_{i=1}^{n} D_{Xi}/n$ shows the proportion of forecasts that correctly identify the direction of change in the actual series.

5 Empirical Results

Table 5 presents the empirical results from the univariate forecasting exercise. The first observation is that no single model is able to provide the best forecast of inbound tourism for all European countries across both the short and long run. Secondly, ETS, NN and ARFIMA models are unable to report the best forecast for any of the countries, at least on one occasion and thus we are able to rule out these models as irrelevant for forecasting European tourism demand. ARIMA, TBATS, and the two SSA models appear to be lucrative based on the RMSE criterion. The findings pertaining to the performance of NN and ETS model forecasts are consistent with the findings in Hassani et al. (2015) where the same two models were seen providing the least favourable forecasts for U.S. tourism demand forecasting. The fact that TBATS reports a better performance than NN and ETS in this case was expected as by definition TBATS was developed for handling time series with complex seasonal patterns (De Livera et al., 2011) and this application shows it is able to report a reasonable performance whilst there is ample room for improvements to this algorithm.

[Insert Table 5 around here]

Table 5 shows in bold font the model with the lowest RMSE at each horizon. Overall, based on the the highest number of bold outcomes reported by a particular model we can suggest that on average across all horizons the two SSA models are able to provide the optimal univariate forecasts in comparison to forecasts from the other models. More specifically, if one is interested in using a single model which can provide the most accurate forecast of tourism demand for a particular country, then we can make the following suggestions. When forecasting tourism demand in Germany, Greece, Cyprus, Portugal, Sweden, and UK the SSA-R model can provide the best forecasts whilst for Italy, Netherlands and Austria SSA-V model is the best option. For Spain the traditional ARIMA model is seen providing the best forecasts on average. Furthermore, focusing on the forecasts with the lowest RMSE at each horizon, then we maintain that this depends on a mixture of forecasting models for a given country based on the horizon of interest.

Let us consider the forecasting results for each country at each horizon in detail. We find that ARIMA provides the best forecasts for tourist arrivals in Germany and Greece in both the very short and very long run whilst SSA-R forecasts outperform the rest at $h = 3, 6, 12$ steps-ahead. For tourist arrivals in Spain, forecasts from TBATS are found to be best at $h = 1, 3, 24$ steps-ahead whilst ARIMA forecasts are seen reporting the lowest RMSE at $h = 6, 12$ steps-ahead. SSA-V forecasts provide the lowest error for tourist arrivals in Italy at horizons of 1, 12 and 24 steps ahead with SSA-R reporting the best forecast at 3 months ahead and ARIMA reporting the best forecast at 6 months ahead. For tourist arrivals in Cyprus, SSA-V forecasts are best in the very short run and SSA-R forecasts are best in the medium term ($h = 3, 6$) whilst TBATS can provide the best forecasts in the long run ($h = 12, 24$). TBATS is seen reporting the best forecast for the Netherlands at $h = 1$ step ahead with SSA-V providing the best forecast at all other horizons. When forecasting tourist arrivals in Austria we find that SSA-V can provide the best forecasts at horizons of 1, 3, 12 and 24 steps-ahead with SSA-R providing the best forecast at $h = 6$ months ahead. For Portugal, ARIMA can provide the best forecast in the very short run whilst SSA-R is best at providing the better forecasts at all remaining horizons. For Sweden, forecasts from ARIMA are best at horizons of 1 and 6 steps ahead whilst SSA-V forecasts are best at 3 and 12 steps ahead with SSA-R providing the best forecast in the very long run. For UK once again ARIMA provides the best forecast in the very short run whilst SSA-R provides the best forecasts for the remaining horizons.

The results in Table 5 also make it clear that the SSA models appear to be best especially beyond $h = 1$ step-ahead as majority of the instances whereby SSA outperforms the other models are in the medium - long term cases. These results are useful to practitioners for various reasons. First, it enables them to easily determine which model is best in general overall for modelling and forecasting tourist arrivals in these selected countries should one only wish to use a single model. Second, the results also enable practitioners to select which model is on average best for forecasting a particular horizon across all countries. Third, a more closer look enables practitioners to pick the best model for forecasting a chosen horizon for each individual country.

However, relying on the RMSE alone for determining the best forecasting model is not statistically efficient. As such, we go a step further and test all our out-of-sample forecasting results for statistical significance using the modified Diebold-Mariano test in Harvey et al. (1997). For this purpose we consider SSA-R forecasts as a benchmark and calculate the RMSE comparing forecasts from each other model against our chosen benchmark. The choice of SSA-R as the benchmark model is a result of many positive aspects. First, for the 10 countries considered in this study, forecasts from the SSA-R model report the lowest average RMSE across all horizons in 6 out of the 10 cases which is equivalent to 60% of all cases. Second, SSA-R forecasts report the highest number of the lowest RMSEs at each horizon for all countries considered here. Thus, based on the criterion of a loss function it is clear that in general the SSA-R is the best performing model overall. However, instead of relying on the RMSE criterion alone, we also consider the Model Confidence Set (MCS) of Hansen et al. (2011). The results show that across all horizons, the SSA-R model is constantly ranked as either first, second or third in comparison to the other nine models which provides added justification for its choice as a benchmark in this study[1].

The RMSE results are reported in Table 6. The RMSE criterion can provide us with the following information. Suppose that we wish to quantify how well a particular model fares against the benchmark, then if we consider the average RMSE between SSA-R and ARIMA forecasts for Germany, the value of 0.93 indicates that SSA-R forecasts are 7% better on average across all horizons than the ARIMA forecasts for same country. In terms of statistically significant differences between forecasts, all forecasts from NN and ARFIMA models are found to have a

[1]The detailed results from Hansen et al.'s (2011) MCS test are available upon request.

statistically significant difference in comparison to forecasts from SSA-R. The score indicates the number of statistically significant outcomes reported by SSA-R in comparison to other models for each country. The percentage score indicates that across all countries the number of statistically significant outcomes have always been at or above a minimum of 50% and thus indicates the results do represent a considerable amount of statistically significant outcomes in this study.

[Insert Table 6 around here]

In line with good statistical practice we also consider the direction of change (DC) predictions of all forecasts. These are reported in Table 7. SSA-V forecasts interestingly reports the largest number of highest average DC predictions across all horizons for the countries considered in this study, whilst ARIMA is second best. The DC results for the SSA-R forecasts are not the best but it is important to remember that the DC criterion should always be coupled with results from a loss function for the accuracy of forecasts in order to make meaningful decisions. Practitioners can use the information in the DC table in combination with the results in Tables 5 and 6 to determine which model to use to obtain forecasts for a particular country based on the objective of the exercise. This enables them to reach a compromise between the accuracy of forecasts in terms of the lowest possible error and the best DC prediction.

[Insert Table 7 around here]

Finally, we go a step further and calculate the cumulative distribution functions (cdf) of the absolute values of the out-of-sample errors for all ten countries across all horizons from all seven forecasting models. According to Hassani et al. (2009), if the cdf graph produced by one method is strictly above the graph of another cdf, then we can conclude that the errors associated with the first method are stochastically smaller than the errors of the second method. In Figure 3 we present a selection of the graphs for instances where there is sufficient evidence to prove that forecasts from SSA-R for all ten countries provide stochastically smaller errors than forecasts from a competing model at a given horizon. Based on the cdf graphs we find concrete evidence to prove that for all ten countries considered in this study, at horizons of 1, 3, 6, 12 and 24 step-ahead-forecasts from the SSA-R suggest stochastically smaller errors than those from forecasts based on the ARFIMA and NN models, as the absolute errors from SSA-R forecasts are lying strictly and well above those of ARFIMA and NN. Furthermore, at $h = 12$ steps-ahead we find comparatively weaker evidence to justify that SSA-R provides stochastically smaller errors than forecasts from ETS and ARIMA at this horizon across all ten countries. The evidence is more convincing at $h = 24$ steps ahead where we find that SSA-R forecasts across all 10 countries at this horizon indicate stochastically smaller errors than those from ETS forecasts. These results provide added justification to the claims based on the RMSE, RRMSE and the MCS test of Hansen et al. (2011) which suggested that SSA-R is the an appropriate benchmark for this study.

[Insert Figure 3 around here]

6 Conclusion

The aim of this paper is to generate and evaluate international tourist arrival forecasts in selected European countries. We focus on short-, medium- and long-run forecasts using several parametric and nonparametric forecasting techniques. The countries under investigation are

Austria, Cyprus, Germany, Greece, Netherlands, Portugal, Spain, Sweden and the United Kingdom and the study period spans from January 2000 until December 2013. Previous studies mainly compare different specifications of a single model or use a limited number of different classes of models. This study provides the most comprehensive forecasting comparison among several parametric and non-parametric techniques, namely, the ARIMA, ETS, NN, TBATS, ARFIMA, SSA-R and SSA-V. Furthermore, this is the first study to use the TBATS and SSA-R models for tourist arrival forecasting purposes.

The results suggest that there is not a single model that its forecasting accuracy consistently outperforms that of all other models for any of the countries under investigation and any of the forecasting horizons. More specifically, based on the RMSE, DC and DM tests, the SSA-R, SSA-V, ARIMA and TBATS models are found to be viable options for modelling European tourist arrivals based on the number of times that they outperform the competing models. Forecasts from NN, ETS and ARFIMA models provide the least accurate predictions for European tourist arrivals.

Overall, these results enable forecasters to choose the most suitable model, based on the country, forecast horizon and direction of change criteria, for forecasting tourism demand. Should a single model be of interest, then, across all selected countries and horizons the SSA-R model is found to be the most efficient based on lowest overall forecasting error.

As previously, future research could be aimed at revisiting the robustness of our results in multivariate nonlinear frameworks, which controls for additional exogenous variables that affect tourism demand. Moreover, an avenue for future research is to examine whether a combination of forecasts based on the aforementioned models provides any additional gains in the forecasting accuracy of tourism demand.

References

Antonakakis, N., Dragouni, M., Filis, G., 2015a. How strong is the linkage between tourism and economic growth in europe? *Economic Modelling* 44, 142–155.

Antonakakis, N., Dragouni, M., Filis, G., 2015b. Tourism and growth: The times they are a-changing. *Annals of Tourism Research* 50, 165–169.

Arslanturk, Y., Balcilar, M., Ozdemir, Z. A., 2011. Time-varying linkages between tourism receipts and economic growth in a small open economy. *Economic Modelling* 28 (1-2), 664–671.

Balcilar, M., van Eyden, R., Inglesi-Lotz, R., Gupta, R., 2014. Time-varying linkages between tourism receipts and economic growth in South Africa. *Applied Economics* 46 (36), 4381–4398.

Brida, J. G., Risso, W. A., 2011. Research note: Tourism demand forecasting with SARIMA models the case of South Tyrol. *Tourism Economics* 17 (1), 209–221.

Burger, C., Dohnal, M., Kathrada, M., Law, R., 2001. A practitioners guide to time-series methods for tourism demand forecasting a case study of Durban, South Africa. *Tourism Management* 22 (4), 403–409.

Chan, Y.-M., 1993. Forecasting tourism: A sine wave time series regression approach. *Journal of Travel Research* 32, 58–60.

Cho, V., 2003. A comparison of three different approaches to tourist arrival forecasting. *Tourism Management* 24 (3), 323 – 330.

Chu, F.-L., 1998a. Forecasting tourism arrivals: A combined approach. *Tourism Management* 19 (6), 515520.

Chu, F.-L., 1998b. Forecasting tourist arrivals: Nonlinear sine wave or ARIMA? *Journal of Travel Research* 36, 79–84.

Chu, F.-L., 2004. Forecasting tourism demand: A cubic polynomial approach. *Tourism Management* 25 (2), 79–88.

Chu, F.-L., 2008. A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism Management* 29 (1), 79–88.

Chu, F.-L., 2009. Forecasting tourism demand with ARMA-based methods. *Tourism Management* 30 (5), 740–751.

Chu, F.-L., 2011. A piecewise linear approach to modeling and forecasting demand for Macau tourism. *Tourism Management* 36 (2), 1414–1420.

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

Coshall, J. T., 2009. Combining volatility and smoothing forecasts of {UK} demand for international tourism. *Tourism Management* 30 (4), 495–511.

De Livera, A. M., Hyndman, R. J., Snyder, R. D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association* 106 (496), 1513–1527.

Dharmaratne, G. S., 1995. Forecasting tourist arrivals in {B}arbados. *Annals of Tourism Research* 22 (4), 804–818.

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

Golyandina, N., Nekrutkin, V., Zhigljavsky, A., 2001. *Analysis of Time Series Structure: SSA and Related Techniques.* New York-London: Chapman and Hall/CRC.

Hadavandi, E., Ghanbari, A., Shahanaghi, K., Abbasian-Naghneh, S., 2011. Tourist arrival forecasting by evolutionary fuzzy systems. *Tourism Management* 32 (5), 1196–1203.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. *Econometrica* 79 (2), 453–497.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13 (2), 281291.

Hassani, H., 2007. Singular spectrum analysis: Methodology and comparison. *Journal of Data Science* 5, 239–257.

Hassani, H., Silva, E. S., Gupta, R., Segnon, M. K., 2015. Forecasting the price of gold. *Applied Economics* 47 (39), 4141–4152.

Hassani, H., Soofi, A. S., Zhigljavsky, A., 2013. Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (3), 743–760.

Kim, J. H., Moosa, I. A., 2005. Forecasting international tourist flows to Australia: A comparison between the direct and indirect methods. *Tourism Management* 26 (1), 69–78.

Kim, J. H., Wong, K., Athanasopoulos, G., Liu, S., 2011. Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals. *International Journal of Forecasting* 27 (3), 887–901.

Kon, S. C., Turner, L. W., 2005. Neural network forecasting of tourism demand. *Tourism Economics* 11 (3), 301–328.

Kulendran, N., Witt, S. F., 2003. Forecasting the demand for international business tourism. *Journal of Travel Research* 41 (3), 265–271.

Law, R., Au, N., 1999. A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management* 20 (1), 89–97.

Lim, C., McAleer, M., 2001. Forecasting tourist arrivals. *Annals of Tourism Research* 28 (4), 965–977.

Pai, P.-F., Hung, K.-C., Lin, K.-P., 2014. Tourism demand forecasting using novel hybrid system. *Expert Systems with Applications* 41 (8), 3691–3702.

Palmer, A., Montao, J. J., Sesé, A., 2006. Designing an artificial neural network for forecasting tourism time series. *Tourism Management* 27 (5), 781–790.

Silva, E. S., Rajapaksa, C. R., 2014. Evaluating the effectiveness of parametric and nonparametric energy consumption forecasts for a developing country. *International Journal of Energy and Statistics* 02 (02), 89–101.

Song, H., Witt, S. F., 2006. Forecasting international tourist flows to Macau. *Tourism Management* 27 (2), 214–224.

Song, H., Witt, S. F., Jensen, T. C., 2003. Tourism forecasting: Accuracy of alternative econometric models. *International Journal of Forecasting* 19 (1), 123–141.

UNWTO, 2014. *UNWTO tourism highlights 2014 edition*. Madrid: UNWTO.

Vu, J. C., Turner, L. W., 2006. Regional data forecasting accuracy: The case of Thailand. *Journal of Travel Research* 45 (2), 186–193.

Wan, S.-K., Wang, S.-H., Woo, C.-K., 2013. Aggregate vs. disaggregate forecast: Case of Hong Kong. *Annals of Tourism Research* 42 (0), 434–438.

Wong, K. K., Song, H., Chon, K. S., 2006. Bayesian models for tourism demand forecasting. *Tourism Management* 27 (5), 773–780.

Wong, K. K., Song, H., Witt, S. F., Wu, D. C., 2007. Tourism forecasting: To combine or not to combine? *Tourism Management* 28 (4), 1068–1078.

Zietz, J., Traian, A., 2014. When was the U.S. housing downturn predictable? A comparison of univariate forecasting methods. *The Quarterly Review of Economics and Finance* 54 (2), 271–281.

Table 1: Descriptive statistics for European tourist arrivals (Jan. 2000 - Dec. 2013).

	Min.	Max.	Mean	Med.	IQR	SD	CV	SW (p)	OCSB
Germany	878100	3895000	1953000	1849000	909447	641550	32	<0.01	1
Greece	88040	2838000	762900	555100	1065914	686634	89	<0.01	1
Spain	1566000	6744000	3449000	3429000	2449991	1366387	39	<0.01	1
Italy	1067000	7457000	3415000	3442000	2863705	1653644	48	<0.01	1
Cyprus	30746	321844	158692	187800	154268	83409	53	<0.01	1
Netherlands	451200	1541087	874767	895900	408633	241150	28	<0.01	0
Austria	492255	2834741	1501201	1480754	661226	475432	32	0.67*	1
Portugal	192923	1181643	533720	531457	369484	227719	43	<0.01	1
Sweden	125916	1428207	383473	240430	187245	303655	79	<0.01	1
United Kingdom	692120	3162159	1628266	1495147	770656	546790	34	<0.01	1

Note: * indicates data is normally distributed based on a Shapiro-Wilk (SW) test at p=0.05. 0 indicates there is no seasonal unit root based on the OCSB test at p=0.05. 1 indicates there is a seasonal unit root based on the OCSB test at p=0.05.

Table 2: Break points in European tourist arrivals series.

Series	Structural Break
Germany	2005(4), 2011(4)
Greece	2009(4)
Spain	2006(3)
Italy	2010(4)
Cyprus	None
Netherlands	2011(3)
Austria	2007(5)
Portugal	2006(3)
Sweden	None
United Kingdom	2005(4)

Table 3: Forecasting model parameters for European tourist arrivals.

Series	ARIMA	ETS(α, γ, σ)	NN(p, P, k)	TBATS	ARFIMA(d)	SSA-V	SSA-R
Germany	(0,1,1)(1,1,1)	(0.485,1e-04,0.0415)^M	NNAR(2,1,1)	(0.357,{0,0},1,{<12,5>})	0.33	(48,13)	(48,13)
Greece	(3,1,1)(0,1,2)	(0.7496,1e-04,0.0796)^M	NNAR(2,1,1)	(0,{0,0},-,{<12,5>})	0.27	(36,20)	(36,20)
Spain	(1,0,2)(0,1,1)	(0.4911,4e-04,0.0308)^M	NNAR(2,1,1)	(0.082,{0,0},-,{<12,5>}	0.27	(36,22)	(36,22)
Italy	(0,0,2)(0,1,1)	(0.2463,1e-04,0.0573)^M	NNAR(2,1,1)	(0,{0,0},0.999,{<12,5>})	0.35	(60,16)	(60,16)
Cyprus	(1,0,1)(2,0,0)	(0.649,0.0015,0.0937)^M	NNAR(2,1,1)	(0.301,{0,0},1,{<12,5>})	0.29	(36,14)	(36,14)
Netherlands	(1,0,2)(2,1,2)*	(0.3069,1e-04,0.0565)^M	NNAR(2,1,1)	(1,{2,0},-,{<12,5>})	0.32	(36,11)	(36,11)
Austria	(2,0,3)(2,1,2)	(0.0628,1e-04,0.0628)^M	NNAR(2,1,1)	(0.263,{1,0},1,{<12,5>})	0.19	(60,20)	(60,20)
Portugal	(1,0,1)(0,1,2)	(0.4834,1e-04,0.0531)^M	NNAR(2,1,1)	(0.027,{0,0},-,{<12,5>})	0.19	(36,14)	(36,14)
Sweden	(1,0,1)(1,1,1)	(0.8362,1e-04,0.0884)^M	NNAR(2,1,1)	(0,{2,0},0.999,{<12,5>})	0.09	(60,20)	(48,20)
United Kingdom	(1,0,3)(0,1,1)	(0.3707,1e-04,127846.3)^M	NNAR(2,1,1)	(0.557,{0,0},-,{<12,4>})	4.58e-05	(48,8)	(48,8)

Note:* indicates an ARIMA model with drift. ^{M} is an ETS model with multiplicative seasonality. α, γ, σ are the ETS smoothing parameters. p is the number of lagged inputs, P is the automatically selected value for seasonal time series, and k is the number of nodes in the hidden layer. d is the differencing parameter. L is the window length and r is the number of eigenvalues.

Table 4: W-correlations between signal and residuals for European arrivals.

Series	SSA-V	SSA-R
Germany	0.005	0.005
Greece	0.006	0.006
Spain	0.005	0.005
Italy	0.004	0.004
Cyprus	0.010	0.010
Netherlands	0.009	0.009
Austria	0.005	0.006
Portugal	0.006	0.006
Sweden	0.020	0.020
United Kingdom	0.014	0.014

Table 5: Out-of-sample RMSE results for European tourist arrivals.

h	Germany	Greece	Spain	Italy	Cyprus	Netherlands	Austria	Portugal	Sweden	UK
ARIMA										
1	**60878**	**87469**	178075	286925	14583	73122	124291	**43163**	**26310**	**169626**
3	83511	162309	264844	300433	26090	82915	120614	64763	26216	236219
6	100444	193335	**283152**	**246547**	28481	87697	112352	75174	**23456**	274589
12	69436	237010	**205741**	249131	23899	111354	134475	76410	29027	314892
24	**67705**	**148597**	338420	392398	30902	183056	160058	124777	50192	326923
Avg.	76395	165744	**254047**	295087	24791	107629	130358	76857	31040	264450
ETS										
1	72512	167965	175616	323764	17224	70263	101502	47590	68145	182703
3	91323	404044	283443	374324	34538	73491	99931	68241	135865	252056
6	110840	511331	371152	305881	43624	85819	101770	94031	71223	281880
12	158631	220320	185833	287974	15988	98392	125114	75547	26005	320905
24	307251	203286	292574	501241	21318	158859	194347	128418	36960	335751
Avg.	148111	301389	261723	358637	26538	97365	124533	82765	67640	274659
NN										
1	583107	927936	1216291	1656221	136427	302251	818291	286229	277131	647598
3	601307	1103090	1331447	1797720	148617	289456	880077	249931	277702	666412
6	592265	791876	1523879	2060446	170584	253952	828686	252160	260963	576560
12	870987	857915	1416616	2384578	191839	314958	562093	289237	241766	526148
24	1043841	947666	2101843	2852818	114903	366222	533021	377186	340947	691906
Avg.	738302	925697	1518015	2150357	152474	305368	724434	290949	279702	621725
TBATS										
1	69755	172827	**165087**	341700	19861	**67244**	112262	52697	59588	185346
3	74674	341103	**251317**	355806	30749	75087	109237	63350	118345	236517
6	82177	460453	339932	277413	38975	83379	108682	94370	82045	262771
12	68885	327041	226438	292928	**17494**	89928	124210	76427	50967	323241
24	96389	413543	**308376**	420722	**25576**	138154	157441	124781	80656	308798
Avg.	78376	342993	258230	337714	26531	90758	122366	82325	78320	263335
ARFIMA										
1	288078	288527	561791	762573	32801	151566	277166	121087	147602	243738
3	498755	615426	1326277	1147309	53293	185584	327985	234275	170292	405534
6	485296	612500	1429433	1058050	51604	228567	325631	231540	188204	437942
12	549514	464374	1551741	764317	52636	187482	358598	206498	191729	426249
24	836800	723297	1931975	1264606	78506	258383	429657	335729	250608	846006
Avg.	531689	540825	1360244	999371	53768	202317	343808	225826	189687	471894
SSA-V										
1	66754	90421	187053	**269987**	**12452**	79624	**110669**	43899	30150	219075
3	74657	161642	258974	247843	19309	**72825**	**97258**	53920	**23912**	238796
6	84512	174413	319133	247302	25103	**76307**	100631	65085	24282	244749
12	71809	198750	273338	**230185**	27319	**82871**	**92983**	75607	**26318**	226785
24	79860	156307	474835	**289067**	36938	**114085**	**131438**	102602	49615	322121
Avg.	75518	156307	302667	**256877**	24224	**85142**	**106596**	68223	30855	250305
SSA-R										
1	66278	87807	197206	273286	14132	80036	111899	45565	27880	214170
3	**69996**	**151700**	343912	**243894**	**17722**	75227	104308	**52128**	24230	**225354**
6	**74054**	**157592**	396543	248764	**17923**	78755	**99183**	**58928**	25766	**232604**
12	**53384**	**173704**	227745	256370	21241	86821	96732	**68175**	26986	**222035**
24	82974	177537	409057	298978	33875	126767	167380	**97214**	**34386**	**280366**
Avg.	**69337**	**149668**	314892	264258	**20979**	89521	115900	**64402**	**27850**	**234906**

Note: Shown in bold font is the model reporting the lowest RMSE at each horizon for a given country.

Table 6: Out-of-sample RMSE results for European tourist arrivals with SSA-R as the benchmark model.

h	Germany	Greece	Spain	Italy	Cyprus	Netherlands	Austria	Portugal	Sweden	UK
$\frac{SSA-R}{ARIMA}$										
1	1.09	1.00	1.11	0.95	0.97	1.09	0.90	1.06	1.06	1.26*
3	0.84	0.93	1.30*	0.81	0.68	0.91	0.86	0.80*	0.92	0.95
6	0.74	0.82	1.40*	1.01	0.63*	0.90	0.88	0.78*	1.10*	0.85
12	0.77	0.73*	1.11	1.03	0.89	0.78	0.72*	0.89*	0.93	0.71
24	1.23*	1.19*	1.21*	0.76*	1.10*	0.69*	1.05*	0.78*	0.69*	0.86*
Avg.	0.93	0.94	1.22	0.91	0.85	0.87	0.88	0.86	0.94	0.93
$\frac{SSA-R}{ETS}$										
1	0.91	0.52*	1.12	0.84	0.82	1.14	1.10	0.96	0.41*	1.17
3	0.77	0.38	1.21	0.65	0.51*	1.02	1.04	0.76	0.18*	0.89
6	0.67*	0.31*	1.07	0.81	0.41*	0.92	0.97	0.63*	0.36*	0.83
12	0.77*	0.73	1.11	1.03	0.89	0.78	0.72	0.89*	0.93	0.71
24	0.27*	0.87*	1.40*	0.60*	1.59*	0.80*	0.86*	0.76*	0.93*	0.84*
Avg.	0.68	0.56	1.18	0.79	0.84	0.93	0.94	0.80	0.56	0.89
$\frac{SSA-R}{NN}$										
1	0.11*	0.09*	0.16*	0.17*	0.10*	0.26*	0.14*	0.16*	0.10*	0.33*
3	0.12*	0.14*	0.26*	0.14*	0.12*	0.26*	0.12*	0.21*	0.09*	0.34*
6	0.13*	0.20*	0.26*	0.12*	0.11*	0.31*	0.12*	0.23*	0.10*	0.40*
12	0.06*	0.20*	0.16*	0.11*	0.11*	0.28*	0.17*	0.24*	0.11*	0.42*
24	0.08*	0.19*	0.19*	0.10*	0.29*	0.35*	0.31*	0.26*	0.10*	0.41*
Avg.	0.10	0.16	0.21	0.13	0.15	0.29	0.17	0.22	0.10	0.38
$\frac{SSA-R}{TBATS}$										
1	0.95	0.51*	1.19	0.80	0.71*	1.19	1.00	0.86	0.47*	1.16
3	0.94	0.44	1.37	0.69	0.58	1.00	0.95	0.82	0.20*	0.95
6	0.90	0.34*	1.17	0.90	0.46*	0.94	0.91	0.62*	0.31*	0.89
12	0.77	0.53	1.01	0.88*	1.21	0.97	0.78	0.89*	0.53	0.69
24	0.86*	0.43*	1.33*	0.71*	1.32*	0.92*	1.06*	0.78*	0.43*	0.91*
Avg.	0.88	0.45	1.21	0.79	0.86	1.00	0.94	0.80	0.39	0.92
$\frac{SSA-R}{ARFIMA}$										
1	0.23*	0.30*	0.35*	0.36*	0.43*	0.53*	0.40*	0.38*	0.19*	0.88*
3	0.14*	0.25*	0.26*	0.21*	0.33*	0.41*	0.32*	0.22*	0.14*	0.56*
6	0.15*	0.26*	0.28*	0.24*	0.35*	0.34*	0.30*	0.25*	0.14*	0.53*
12	0.10*	0.37*	0.15*	0.34*	0.40*	0.46*	0.27*	0.33*	0.14*	0.52*
24	0.10*	0.25*	0.21*	0.24*	0.43*	0.49*	0.39*	0.29*	0.14*	0.33*
Avg.	0.14	0.29	0.25	0.28	0.39	0.45	0.34	0.29	0.15	0.56
$\frac{SSA-R}{SSA-V}$										
1	0.99	0.97	1.05	1.01	1.13*	1.01	1.01	1.04	0.92	0.98
3	0.94	0.94	1.33	0.98	0.92	1.03	1.07	0.97	1.01	0.94
6	0.88	0.90	1.24*	1.01	0.71*	1.03*	0.99	0.91*	1.06	0.95
12	0.74*	0.87	0.83*	1.11	0.78*	1.05*	1.04*	0.90	1.03	0.98*
24	1.04*	1.14*	0.86*	1.03*	0.92*	1.11*	1.27*	0.95*	0.69*	0.87*
Avg.	0.92	0.96	1.06	1.03	0.89	1.05	1.08	0.95	0.94	0.94
Score	17	19	18	15	22	16	16	22	21	16
% Score	0.56	0.63	0.60	0.50	0.73	0.53	0.53	0.73	0.70	0.53

Note: * indicates a statistically significant difference between forecasts based on the modified Diebold Mariano test at $p = 0.10$. Score indicates the number of statistically significant outcomes for each horizon.

Table 7: Direction of change results for European tourist arrivals.

h	Germany	Greece	Spain	Italy	Cyprus	Netherlands	Austria	Portugal	Sweden	UK
ARIMA										
1	1.00*	0.91*	1.00*	0.97*	0.91*	0.84*	0.84*	1.00*	0.94*	0.84*
3	1.00*	0.90*	1.00*	1.00*	0.97*	0.97*	0.97*	1.00*	0.97*	0.93*
6	1.00*	1.00*	1.00*	0.89*	1.00*	1.00*	0.81*	0.96*	1.00*	0.85*
12	1.00*	0.38	0.43	0.52	0.38	0.33	0.43	0.29	0.62	0.57
24	1.00*	0.89*	0.33	1.00*	0.22	0.44	0.67	0.56	0.44	0.33
Avg.	**1.00**	**0.82**	0.75	**0.88**	0.70	0.72	0.74	0.76	0.79	0.71
ETS										
1	0.91*	0.91*	0.94*	1.00*	0.91*	0.91*	0.94*	1.00*	0.94*	0.91*
3	1.00*	0.97*	1.00*	1.00*	1.00*	1.00*	0.97*	1.00*	0.97*	0.90*
6	1.00*	1.00*	1.00*	0.93*	1.00*	1.00*	0.93*	0.93*	1.00*	0.85*
12	0.24	0.52	0.57	0.52	0.19	0.67	0.52	0.38	0.67	0.48
24	0.22	0.56	0.56	0.67	0.22	0.67	0.56	0.33	0.33	0.44
Avg.	0.67	0.79	**0.81**	0.82	0.66	0.85	0.78	0.73	0.78	0.72
NN										
1	0.91*	0.47	0.75*	0.81*	0.72*	0.63	0.34	0.56	0.59	0.41
3	0.77*	0.77*	0.77*	0.80*	0.67	0.80*	0.70*	0.73*	1.00*	0.83*
6	0.81*	0.85*	0.78*	0.85*	0.70*	0.85*	0.52	0.93*	0.85*	0.85*
12	0.52	0.48	0.38	0.33	0.19	0.52	0.43	0.38	0.67	0.62
24	0.33	0.44	0.22	0.56	0.11	0.67	0.67	0.22	0.89*	0.56
Avg.	0.67	0.60	0.58	0.67	0.48	0.69	0.53	0.56	0.80	0.65
TBATS										
1	0.91*	0.84*	1.00*	0.97*	0.88*	0.91*	0.88*	1.00*	0.91*	0.91*
3	1.00*	0.97*	1.00*	1.00*	0.97*	0.93*	1.00*	1.00*	1.00*	0.90*
6	1.00*	1.00*	1.00*	0.93*	1.00*	1.00*	0.89*	0.93*	1.00*	0.89*
12	1.00*	0.62	0.43	0.57	0.38	0.86*	0.38	0.38	0.67	0.48
24	1.00*	0.33	0.33	0.78	0.33	0.89*	0.44	0.33	0.56	0.44
Avg.	0.98	0.75	0.75	0.85	0.71	**0.92**	0.72	0.73	0.83	0.72
ARFIMA										
1	0.84*	0.88*	0.84*	0.88*	0.97*	0.75*	0.84*	0.75*	0.78*	0.75*
3	0.77*	0.97*	0.73*	0.80*	0.87*	0.80*	0.83*	0.77*	0.93*	0.83*
6	0.89*	0.93*	0.89*	0.89*	1.00*	0.93*	0.70*	0.89*	0.85*	0.85*
12	0.10	0.57	0.38	0.38	0.29	0.38	0.33	0.33	0.62	0.52
24	0.00	0.11	0.33	0.33	0.11	0.33	0.33	0.22	0.67	0.22
Avg.	0.52	0.69	0.64	0.66	0.65	0.64	0.61	0.59	0.77	0.64
SSA-V										
1	0.97*	0.88*	0.97*	0.94*	0.88*	0.84*	0.91*	0.94*	0.94*	0.84*
3	0.93*	0.93*	1.00*	1.00*	1.00*	0.93*	0.93*	0.97*	0.97*	0.97*
6	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*	0.89*	1.00*	0.85*	0.93*
12	0.90*	0.43	0.48	0.67	0.38	0.81*	0.81*	0.57	0.76	0.76
24	1.00*	0.67	0.11	0.78	0.67	0.78	0.78	0.78	0.78	0.56
Avg.	0.96	0.78	0.71	**0.88**	**0.78**	0.87	**0.86**	**0.85**	**0.86**	**0.81**
SSA-R										
1	0.97*	0.88*	0.88*	0.97*	0.84*	0.81*	0.91*	0.94*	0.94*	0.84*
3	0.60	0.50	0.57	0.47	0.33	0.57	0.60	0.60	0.47	0.63
6	0.67	0.52	0.48	0.44	0.30	0.48	0.48	0.74*	0.41	0.56
12	0.57	0.52	0.86*	0.57	0.33	0.43	0.48	0.81*	0.48	0.67
24	0.78	0.33	1.00*	0.89*	0.22	0.78	0.78	1.00*	0.44	0.89*
Avg.	0.72	0.55	0.76	0.67	0.41	0.61	0.65	0.82	0.55	0.72

Note: Shown in bold font is the model reporting the best average DC prediction across all horizons for a given country. * indicates the DC predictions are statistically significant based on a t-test at $p = 0.05$.

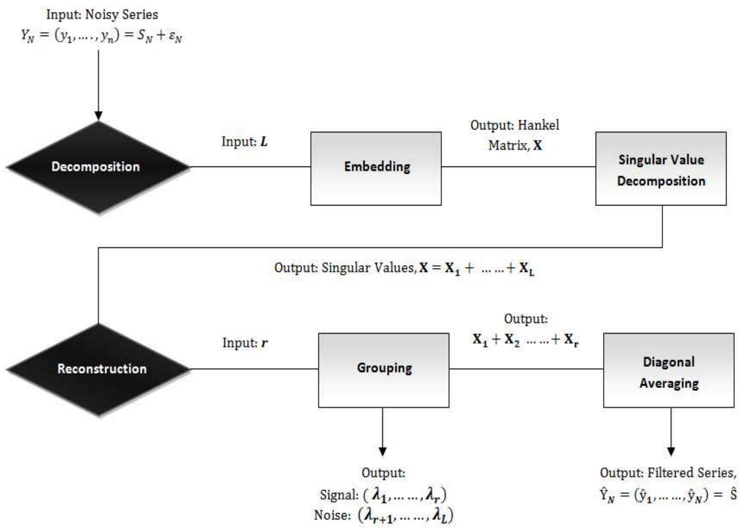

Figure 1: A summary of the basic SSA process.

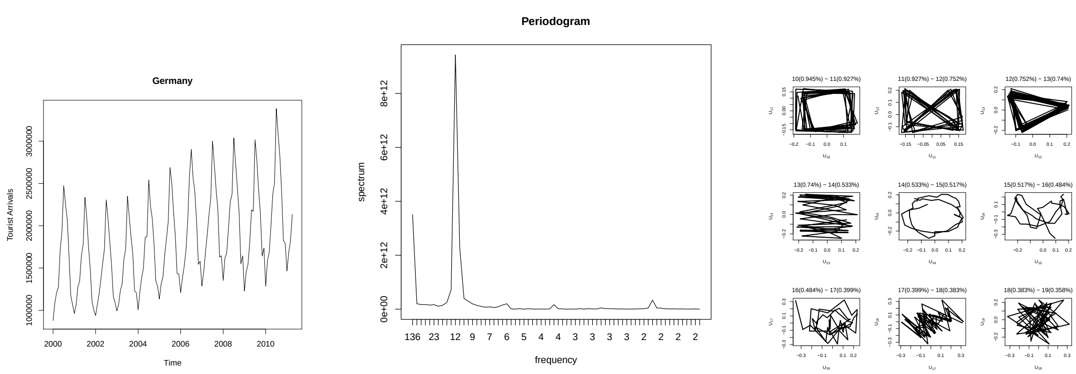

Figure 2: Time series, periodogram and selected paired eigenvectors for German tourist arrivals (Jan. 2000 - Apr. 2011).

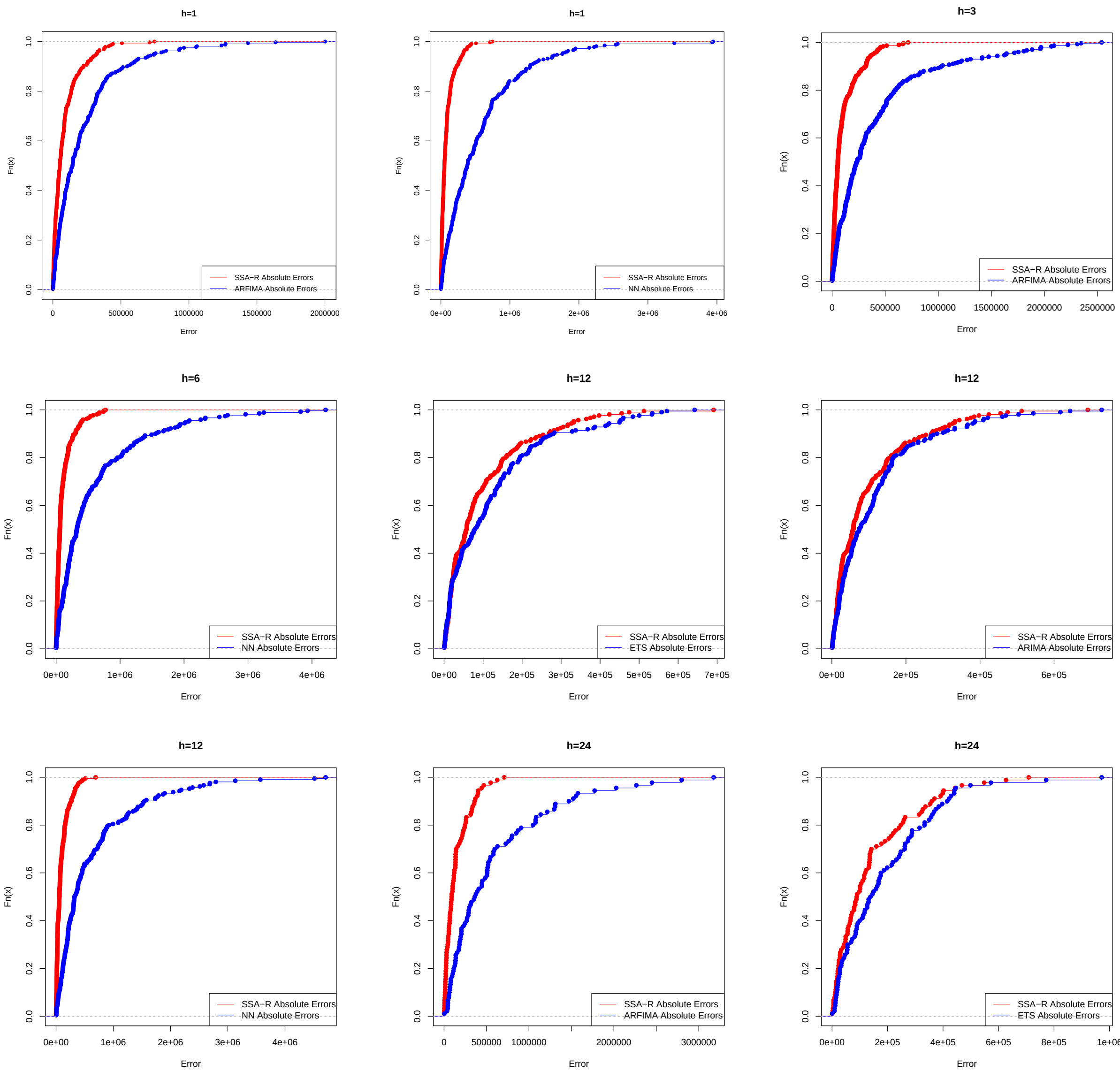

Figure 3: The cumulative distribution functions of the absolute values of the out-of-sample errors.