THEORETICAL ADVANCEMENTS AND APPLICATIONS IN SINGULAR SPECTRUM ANALYSIS

Bournemouth University

AGAMPODIGE EMMANUEL DIYANATH SIRIMAL SILVA

A thesis submitted in partial fulfilment of the requirements of Bournemouth University for the degree of

Doctor of Philosophy

May 2016

Abstract

Singular Spectrum Analysis (SSA) is a nonparametric time series analysis and forecasting technique which has witnessed an augment in applications in the recent past. The increased application of SSA is closely associated with its superior filtering and signal extraction capabilities which also differentiates it from the classical time series methods. In brief, the SSA process consists of decomposing a time series for signal extraction and then reconstructing a less noisy series which is used for forecasting. The aim of this research is to develop theoretical advancements in SSA, supported by empirical evidence to further promote the value, effectiveness and applicability of the technique in the field of time series analysis and forecasting. To that end, this research has four main contributions.

Initially, given the reliance of this research towards improving forecasting processes, it is mandatory to compare and distinguish between the predictive accuracy of forecasts for statistically significant differences. The first contribution of this research is the introduction of a complement statistical test for comparing between the predictive accuracy of two forecasts. The proposed test is based on the principles of cumulative distribution functions and stochastic dominance, and is evaluated via both a simulation study and empirical evidence.

Governments, practitioners, researchers and private organizations publish a variety of forecasts each year. Such forecasts are generally computed using multivariate models and are widely used in decision making processes given the

considerably high level of anticipated forecast accuracy. The classical multi-variate methods consider modelling multiple information pertaining to the same time period or with a time lag into the past. Multivariate Singular Spectrum Analysis (MSSA) is a relatively new and alternative technique for generating forecasts from multiple time series. The second contribution of this research is the introduction of a novel theoretical development which seeks to exploit the information contained in published forecasts (which represent data with a time lag into the future) for generating a new and improved (comparatively more accurate) forecast by taking advantage of the MSSA technique's capability at modelling time series with different series lengths. In brief, the proposed multivariate theoretical development seeks to exploit the forecastability of forecasts by considering not only official and professional forecasts, but also forecasts obtained via other time series models.

The productive application of SSA and MSSA depends largely on the selection of SSA and MSSA parameters, i.e. the Window Length, L, and the number of eigenvalues r which are used for decomposition and reconstruction of time series. Over the years, a variety of mathematically complex, time consuming and labour intensive approaches which require detailed knowledge on the theory underlying SSA have been proposed and developed for the selection of SSA and MSSA parameters. However, the highly labour intensive and complex nature of such approaches have not only discouraged the application of this method by those not conversant with the underlying theory, but also limited SSA and MSSA to offline applications. The third and final contribution of this research proposes new, automated and optimized, SSA and MSSA algorithms for the selection of SSA parameters and thereby enables obtaining optimal SSA and MSSA forecasts (optimized by minimising a loss function). This development opens up the possibility of using SSA and MSSA for online forecasting in the future.

Table of Contents

List of Tables

List of Figures

List of Publications

Hassani, H., Xu, H., Silva, E.S., and Ghodsi, M. (2016). A review of Data Mining applications in Crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, In Press. DOI:10.1002/sam.11312.

Hassani, H., Ghodsi, Z., Silva, E.S., and Heravi, S. (2016). From nature to maths: Improving forecasting performance in subspace-based methods using genetics Colonial Theory. *Digital Signal Processing*, **51**, 101-109.

Hassani, H. and Silva, E.S. (2015). A Kolmogorov-Smirnov based test for comparing between the predictive accuracy of two sets of forecasts. *Econometrics*, **3**(3), 590-609.

Hassani, H., Silva, E.S., and Gupta, R. (2015). Forecasting the Price of Gold. *Applied Economics*, **47**(39), 4141-4152.

Hassani, H., Webster, A., Silva, E.S., and Heravi, S., (2015). Forecasting U.S. Tourist arrivals using optimal Singular Spectrum Analysis. *Tourism Management*, **46**, 322-335.

Silva, E.S., and Hassani, H., (2015). Evaluating the Impact of the 2008 recession on U.S. Trade Forecasting Models. *International Economics*, **141**, 34-49.

Hassani, H., and Silva, E.S., (2015). Forecasting with Big Data: A Review. *Annals of Data Science*, **2**(1), 5-19.

Ghodsi, Z., Silva, E.S., and Hassani, H., (2015). Bicoid Signal Extraction. *Genomics, Proteomics & Bioinformatics*, **13**(3), 183-191.

Hassani, H., Mahmoudvand, R., Omer, H.N., and Silva, E.S. (2014). A Preliminary Investigation into the Effect of Outlier(s) on Singular Spectrum Analysis. *Fluctuation and Noise Letters (FNL)*, **13**(4), 1450029-1-1450029-23.

Hassani, H., Saporta, G., and Silva, E.S. (2014). Data Mining and Official Statistics: The Past, the Present and the Future. *Big Data*, **2**(1), 34-43.

Silva, E.S. and C.R. Rajapaksa (2014). Evaluating the effectiveness of parametric and nonparametric energy consumption forecasts for a developing country. *International Journal of Energy and Statistics*, **2**(2), 89-101.

Silva, E.S., Wu, Y., and Ojiako, U. (2013). Developing Risk Management as a Competitive Capability. *Strategic Change*, **22**(5-6), 281-294.

Silva, E.S. (2013). A combination forecast for energy-related CO2 emissions in the United States. *International Journal of Energy and Statistics*, **1**(4), 269-279.

Beneki, C., and Silva, E.S. (2013). Analysing and forecasting European Union energy data. *International Journal of Energy and Statistics*, **1**(2), 127-141.

Acknowledgements

I would like to dedicate this thesis to my loving Parents and my late Grand Parents.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

AGAMPODIGE EMMANUEL DIYANATH SIRIMAL SILVA

May 2016

List of Abbreviations

ACF Autocorrelation Function

AIC Akaike Information Criterion

AR Autoregressive

ARFIMA Autoregressive Fractionally Integrated Moving Average

ARIMA Autoregressive Integrated Moving Average

ARMA Autoregressive Moving Average

CPI Consumer Price Index

DC Direction of Change

DM Diebold-Mariano

ECM Error Correction Model

EGARCH Exponential GARCH

EIA Energy Information Administration

ESICU.US Industrial Sector Average Regional Electricity Prices in United States

ETS Exponential Smoothing

EXRCP.US Residential Sector Total Electricity Sales in United States

GARCH Generalized Autoregressive Conditional Heteroskedasticity

GDP Gross Domestic Product

HMSSA Horizontal MSSA

HMSSA-R Horizontal MSSA Recurrent forecasting algorithm

HW Holt-Winters

KPSS Kwiatkowski–Phillips–Schmidt–Shin

KS Kolmogorov-Smirnov

KSPA Kolmogorov-Smirnov Predictive Accuracy

LRF Linear Recurrent Formula

MA Moving Average

MAPE Mean Absolute Percentage Error

MARIMA Multivariate ARIMA

MCS Model Confidence Set

MLE Maximum Likelihood Estimator

MSSA Multivariate Singular Spectrum Analysis

Max. Maximum

Min. Minimum

N Length of time series

N/A Not Applicable

NGRCUUS Average Residential Natural Gas Price in United States

NN Neural Networks

OF Official forecast

PF Professional forecast

RMSE Root Mean Squared Error

RRMSE Ratio of the Root Mean Squared Error

SARIMA Seasonal ARIMA

SARS Severe Acute Respiratory Syndrome

SLS Stage Least Squares

SPA Superior Predictive Ability

SSA Singular Spectrum Analysis

SSA-R Recurrent SSA

SSA-V Vector SSA

SVD Singular Value Decomposition

SW Shapiro-Wilk

Skew. Skewness

Std. Dev. Standard Deviation

TBATS ETS with Box–Cox transformation, ARMA errors, trend and seasonal components

TF Transfer Function

TGARCH Threshold GARCH

U.K. United Kingdom

U.S. United States

VAR Vector Autoregression

VFR Visiting friends and relatives

VMSSA Vertical MSSA

VMSSA-R Vertical MSSA Recurrent forecasting algorithm

VMSSA-R Vertical MSSA Recurrent forecasting algorithm

VMSSA-V Vertical MSSA Vector forecasting algorithm

WTI West Texas Intermediary

WTIPUUS West Texas Intermediate Spot Average Crude Oil Price

c.d.f Cumulative Distribution Function

Chapter 1

Introduction

1.1 Time Series Analysis and Forecasting

In a world troubled with ever increasing uncertainty following the on-set of the recent financial crisis in 2008 there is a renewed and opportune demand for methods which can generate improved and accurate forecasts. Such predictions into the future are facilitated via a process known as time series analysis and forecasting. In brief, all time series analysis and forecasting methods can be listed as either parametric or nonparametric techniques. Parametric techniques have the disadvantage of being restricted by assumptions relating to normality and stationarity, whereas nonparametric techniques are model free and are not restricted by any such assumptions. In addition, time series methods can be further classified as univariate and multivariate. Univariate methods consider a single time series whereas multivariate methods consider multiple time series for generating a forecast.

Research and development has led to a wide range of classical (parametric) and novel time series analysis techniques (nonparametric) which include (but are not limited to) Autoregressive Integrated Moving Average (ARIMA), Holt-

Winters (HW), Exponential Smoothing (ETS) and Neural Networks (NN). Each technique has its own benefits and drawbacks, and there is no single method which is identified as universally best at present. This in turn means that the forecasting performance of each technique is largely dependent on both its theoretical underpinning and the nature of the data that is input into the model. As such, different applications will provide different outcomes showing one model outperforming another and vice-versa. Therefore, researchers constantly endeavour to develop more efficient time series analysis models which can provide greater accuracy in comparison to the existing methods.

However, the forecasting models alone are not the only concern for a broad field such as time series analysis. It is clear that when we forecast any variable, there is always an associated loss or error when the forecast is compared with actual data. Various loss functions have been developed to quantify these errors and few which are used in this research are discussed in the chapters which follow. Whilst these loss functions enable comparisons between forecasts from different models, they are unable to determine the statistical significance of these differences. Accordingly, various statistical tests have also been developed over the years to compare between the predictive accuracy of forecasts. In this sense, it is possible to summarize the time series analysis and forecasting process as one which begins with the modelling of data using time series analysis techniques, followed by the generation of forecasts and ending with tests for statistically significant differences between forecasts from competing models.

The gist of this thesis concentrates on a popular nonparametric time series analysis and forecasting technique known as Singular Spectrum Analysis (SSA). The emergence of SSA is closely associated with the work of Broomhead and King (1986a,b) where the authors show that Singular Value Decomposition (SVD) can be used as an effective tool for noise reduction. This was followed by several methodological advancements in SSA and related applica-

tions through Danilov and Zhigljavsky (1997); Fraedrich (1986); Vautard and Ghil (1989); Vautard et al. (1992). In brief, the SSA technique seeks to decompose and filter a time series, and then reconstruct a less noisy series which can be used for forecasting (Hassani, 2007). As such, SSA performs three distinct tasks which can be categorized as decomposition, reconstruction and forecasting. At present, SSA is widely adopted for solving complex issues in the field of time series analysis not only as a forecasting model (Hassani et al., 2009, 2013b), but also as a filtering technique (Hassani et al., 2010a,b). The increasing popularity of SSA is further attributable to its capability of handling both linear, and nonlinear, stationary and non-stationary time series (Hassani, 2007).

A Google Scholar search for applications of SSA shows increased application especially in the new millennium and mostly in the recent past, since 2007 in particular. Accordingly, there is a huge scope for further improving and enhancing the SSA technique as a viable and effective tool for modelling and forecasting in the future. This research study takes advantage of this opportunity and seeks to introduce lucrative theoretical developments for the SSA technique, well supported via empirical evidence.

1.1.1   Research Aim and Objectives

This research is governed by a single aim which is supported by several objectives. These objectives not only enable achieving the aim of the research, but also represents the contributions to SSA and the field of time series analysis and forecasting. The aim of this research is to introduce a complement statistical test for comparing between the predictive accuracy of forecasts and develop theoretical advancements in SSA, supported by empirical evidence to further promote the value, effectiveness and applicability of SSA in the field of time series analysis and forecasting. To that end, this research has four main objectives.

- Introduce a test for comparing the predictive accuracy of two forecasts.

- Develop a new Multivariate SSA (MSSA) based theory for exploiting the forecastability of forecasts.

- Develop a new approach for the selection of SSA and MSSA parameters.

The realization of these objectives will indeed result in considerable theoretical advancements for SSA and also the field of time series analysis and forecasting in general. In addition, emphasis is also placed on providing empirical evidence to portray the practicality of each of the contributions, and some of the applications themselves are the first instances in which the SSA technique is exploited for modelling and forecasting in certain industries.

1.2   Motivation

1.2.1   Why Singular Spectrum Analysis?

Given that the research aims and objectives have been specifically identified above, in this section the objectives are further elaborated upon as the motivation for the selection of SSA, and each of the objectives of this research are concisely explained.

The choice of SSA as the main forecasting tool of interest for this research is motivated by different aspects. Prior to explaining these, it is important to briefly outline the components of a time series. In general, a time series comprises of the signal and noise. As an example, shown in Figure 1.1 is a time series Y which is what we are faced with in reality. However, Y consists of signal and

noise, and as indicated, this signal could be further decomposed into the trend component and sine component in this particular example.

First and foremost, there is a difference in the modelling procedure employed by SSA and classical time series techniques. The classical time series methods consider modelling and forecasting Y. However, the SSA technique will filter Y such that the trend, signal and noise could be identified separately. Thereafter, SSA reconstructs a new time series which corresponds to a less noisy approximation of the signal, for generating forecasts. As such, by using a method such as SSA one is able to obtain a richer understanding of the dynamics underlying a given time series, forecast time series components separately (for example, forecast the trend or seasonal variation alone), and obtain a more accurate overall forecast as the model considers filtering noise which is effectively the random, unexplained components in any given time series.
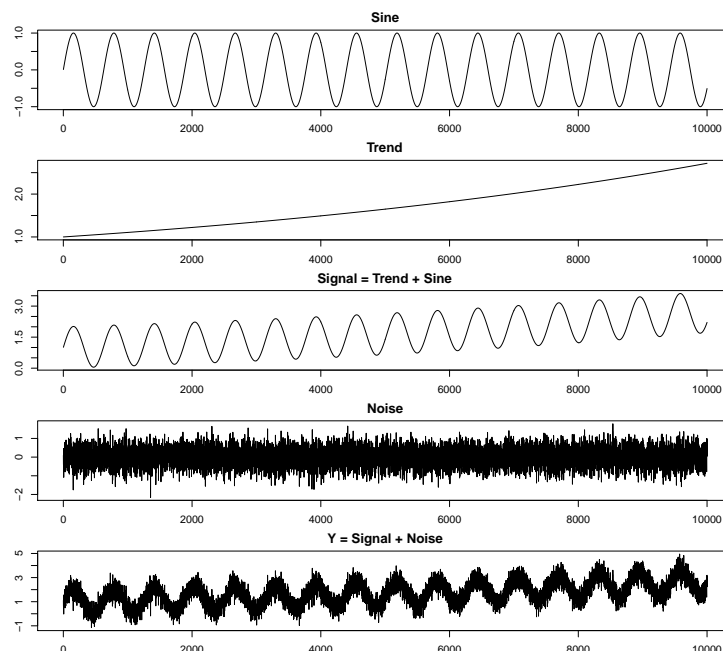
Fig. 1.1 Sine, trend, signal and noise components in Y (Sanei and Hassani, 2015).

Secondly, SSA is a nonparametric technique which does not rely on the as-

sumptions of normality for the residuals, stationarity of the data, and linearity for the model (Hassani et al., 2013a) which are highly unlikely to hold in real world applications. As such, by adopting SSA one is able to model the data sans data transformations which in turn enables a true approximation of the real situation without the loss of any information (Hassani et al., 2013c).

Thirdly, unnatural phenomenal events are known to create outliers in time series data and such outliers in turn result in making a time series non-stationary (Hassani et al., 2014; Tsay, 1998). Given the highly volatile economic conditions experienced in the modern world, it is almost certain that most (economic) time series are affected by the presence of such outliers. Therefore, developing a method such as SSA which is less sensitive to recessions (Hassani et al., 2013a; Silva and Hassani, 2015) can be of added use to future generations.

1.2.2 Why These Objectives?

In general, this research study is primarily aimed at improving a time series analysis and forecasting technique. As such, comparing the predictive accuracy of forecasts generated via different models is not only good statistical practice, but also a mandatory component in ensuring the reliability and validity of the results. At present there exists various statistical tests which are used for comparing between the predictive accuracy of forecasts. See for example, Diebold and Mariano (1995); Hansen (2005); Hansen et al. (2011) and references therein. As the first contribution of this research, a complement statistical test which is founded upon the principles of cumulative distribution functions and stochastic dominance is introduced. The reliability and applicability of the proposed test is evaluated via a simulation study and a corresponding application to real data. Following the successful introduction of a complement statistical test the thesis continues to focus on enhancing SSA and MSSA techniques. It is noteworthy

that for consistency, all applications pertaining to achieving the objectives of this thesis considers tourism data as a common front. This is in addition to the consideration given to data from various other industries in order to portray the applicability and relevance of the proposed approaches in general. Moreover, selected components of the R codes used in this work have been presented via the appendix.

Forecasting is now universal. Practitioners, researchers, professional forecasters and government organizations in particular publish forecasts monthly, quarterly and annually for a variety of variables. Such forecasts are generated via both new and complex univariate and multivariate models which are comparatively more accurate than most classical time series methods. The second objective of this research aims to answer an interesting question, that is, once a forecast is generated, is there a possibility of exploiting this forecast for obtaining a more accurate forecast?

The SSA technique is blessed with both univariate (SSA) and multivariate (MSSA) capabilities. In general, the classical multivariate methods (for example, Vector Autoregression) consider modelling information pertaining to the same time period or with a time lag into the past. In particular, most of the existing multivariate methods can only model and forecast using multiple time series with the same length. However, MSSA has the ability of modelling and forecasting using time series with different series lengths (Hassani and Mahmoudvand, 2013) and this prime advantage in MSSA is used to find a solution to achieve the first objective of this research.

Accordingly, the second contribution of this study is a theoretical development which seeks to exploit the forecastability of forecasts, and thereby promote modelling and forecasting using data with a time lag into the future. The proposed MSSA theoretical development is evaluated for its ability at improving not only existing official and professional forecasts, but also forecasts from

other time series models. It is believed that the results from this research will be of utmost importance to forecasters in general as an innovative and promising research avenue is created in the area of multivariate forecasting.

The third objective of this research is concerned with the selection of SSA and MSSA parameters. This is important because the success of SSA and MSSA techniques depends largely on the accurate selection of its parameters which are referred to as the Window Length (L), and the number of eigenvalues (r) (Sanei and Hassani, 2015). For example, the success of the decomposition stage of SSA and MSSA depends on L whilst the success of the reconstruction and forecasting stages depend on the correct choice of r. Over the years, a variety of mathematically complex, time consuming and labour intensive approaches which require detailed knowledge on the theory underlying SSA have been proposed and developed for the selection of L and r.

Whilst these existing approaches are extremely useful in improving the accuracy of SSA and MSSA functions, they do have two major disadvantages which act as a restriction for the application and use of SSA and MSSA. Firstly, the highly labour intensive nature of the historical approaches for selecting L and r are not only time consuming, but also restricts SSA and MSSA to offline applications. Secondly, the complex and advanced statistical knowledge required to understand the process underlying the selection of L and r in most instances act as a hindrance for the application and use of SSA and MSSA by those not conversant with the advanced statistical theory underlying these techniques.

However, it is important to remember that problems related to complexities surrounding the selection of model parameters in time series analysis and forecasting techniques are universal. As a solution, researchers endeavour to develop automated time series analysis and forecasting methods, and a sound example is the forecast package in R (Hyndman and Khandakar, 2008). Motivated by such efforts, and the interest in promoting the application of SSA, pro-

posed as the third and final contribution of this research is novel, automated and optimized, SSA and MSSA algorithms for the selection of L and r for obtaining optimal SSA and MSSA forecasts (optimized by minimising a loss function). This algorithm for the automation of the SSA and MSSA processes opens up the possibility of using SSA and MSSA for online forecasting.

The remainder of this thesis is organized as follows. Chapter 2 presents SSA and MSSA methodology. Chapter 3 introduces the test for comparing the predictive accuracy of forecasts. Chapter 4 is devoted to the new theoretical development for exploiting the forecastability of forecasts and related empirical applications. Chapter 5 presents the automated and optimized SSA and MSSA algorithms along with empirical evidence and this thesis ends in Chapter 7 along with conclusions, limitations and pathways for future research.

Chapter 2

Methodology

The aim of this Chapter is to introduce the methodology relevant to this study. The main focus is on the theory underlying SSA and MSSA. In addition, the other forecasting techniques and metrics considered for comparative purposes are also briefly explained.

2.1 Singular Spectrum Analysis

The Singular Spectrum Analysis (SSA) technique consists of two stages known as decomposition and reconstruction, each with two complementary steps (Hassani, 2007) which are explained in detail below. In brief, SSA decomposes a time series and thereby enables differentiation between trend, harmonic and noise components, and then reconstructs a less noisy time series using the estimated trend and harmonic components, and this newly reconstructed series is then used to compute forecasts (Golyandina et al., 2001). As a nonparametric method, SSA can be used without making any assumptions pertaining to stationarity and normality of the data (Sanei and Hassani, 2015). This in turn means that no data transformations are required and it is advantageous as the use of parametric techniques would require data transformations in most cases

to ensure the data conforms with the parametric restrictions, and such transformations result in a loss of information (Hassani et al., 2013c).

As a time series method, the SSA technique has both univariate and multivariate capabilities. Applications of univariate SSA for finding solutions to real world problems are diverse, and some recent examples are Ghodsi et al. (2015); Hassani et al. (2010a, 2013a, 2009); Hassani and Thomakos (2010); Rodrıguez-Aragón and Zhigljavsky (2010); Sanei et al. (2011); Silva and Hassani (2015). On the other hand, applications of MSSA are comparatively sparse as MSSA is considered to be relatively new in relation to its univariate counterpart, see for example Groth and Ghil (2011); Hassani et al. (2013b); Hassani and Mahmoudvand (2013); Kapl and Müller (2010); Oropeza and Sacchi (2011); Patterson et al. (2011).

The performance of the SSA technique depends upon the selection of its two parameters known as *i*) the window length L, and *ii*) the number of eigenvalues r. The choice of L and r is discussed in detail in the next chapter. In brief, Sanei and Hassani (2015) notes that the choice of L can vary based on the data one is analysing, the aim of the analysis and the forecasting horizon whilst the incorrect selection of r can result in some parts of the signal(s) being lost, or noise included in the reconstructed series which is effectively made less accurate. In terms of its forecasting capabilities, the SSA technique has two univariate forecasting approaches called recurrent and vector algorithms (Golyandina et al., 2001). In this research, both forecasting approaches for univariate and multivariate SSA are exploited and improved upon.

The entire SSA process can be summarized with the aid of the flow chart in Figure 2.1. According to Sanei and Hassani (2015), initially we are faced with a noisy time series Y_N and the single SSA choice applicable to the decomposition stage, L as inputs. Following a process termed as embedding, we obtain the Hankel matrix $\mathbf{X}$ which is then forwarded as an input into the SVD step. The SVD

step results in singular values which are analyzed to identify and differentiate between signal and noise components. At the reconstruction stage, the singular values are grouped along with the input of the second and final SSA parameter r which results in the grouping matrices $\mathbf{X}_1, \ldots, \mathbf{X}_L$ as either signal or noise. Finally, diagonal averaging is used to to transform the matrices containing signal components into a Hankel matrix so that it can subsequently be converted into a time series which can then be used to forecast future data points.
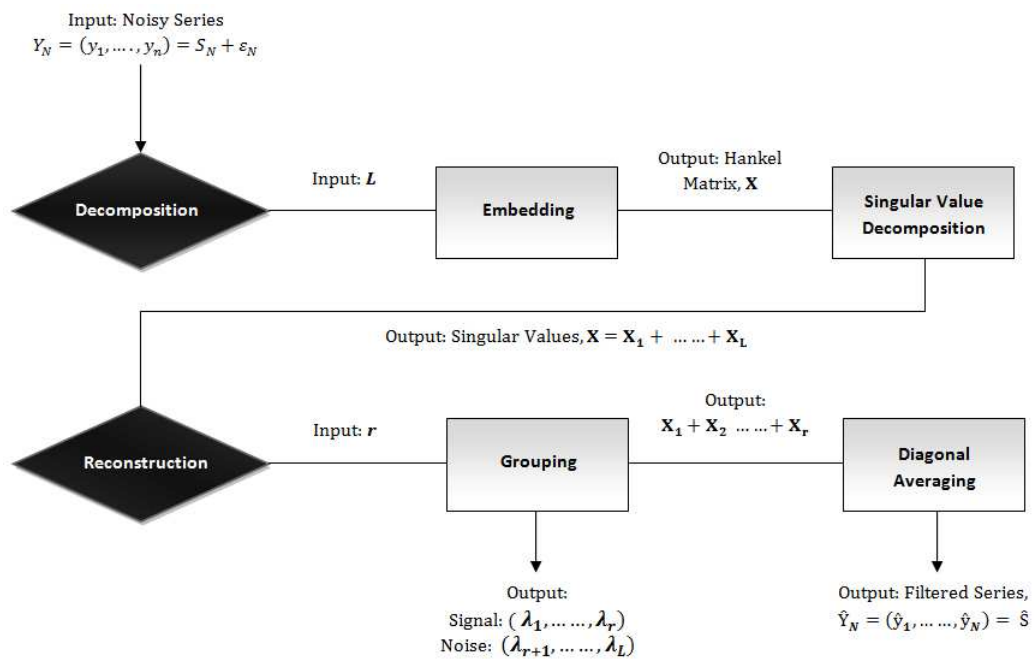
Fig. 2.1 A summary of the basic SSA process (Sanei and Hassani, 2015).

The theory underlying univariate SSA is explained below by following Hassani (2007) and Sanei and Hassani (2015).

2.2   Univariate SSA

Prior to explaining the theory underlying SSA, a simple explanation of the general idea underlying SSA is introduced by following Hassani (2007) and Sanei

and Hassani (2015). Let us consider a noisy time series Y_N with any arbitrary series length N, such that:

$$Y_N = (y_1, \ldots, y_N).$$ (2.1)

Then, let us assume that Y_N comprises of signal and noise. Therefore, Y_N can also be represented as:

$$Y_N = S_N + E_N = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix},$$ (2.2)

where S_N represents the signal and E_N represents noise.

Recall Figure 1.1 where the signal is formed by combining sine and an exponential trend. The classical time series methods will model and forecast Y_N which suggests that such methods consider both the signal and noise in a given series. However, SSA will begin with Y_N, and seek to separate the signal from the noise. Thereafter, it is the filtered, approximated signal that is used to forecast future data points, leaving aside the approximated E_N. Note that the term 'approximated' is used as in practice one is unable to extract the complete signal.

2.2.1 Stage 1: Decomposition

At the decomposition stage, the Window Length L is the only parameter which is relevant as SSA organizes the one dimensional time series Y_N into a multidimensional series. Note that L, is an integer such that $2 \leq L \leq N/1$.

Step 1: Embedding

In the most basic terms, embedding is a mapping operation that transfers a one-dimensional time series Y_N into a multidimensional series $X_1, \ldots, X_K$ with vectors

$$X_i = [y_i, y_{i+1}, y_{i+2}, \ldots, y_{i+L-1}]^T, \tag{2.3}$$

for $i = 1, 2, \ldots, K$ and T denotes transposition. As mentioned in Sanei and Hassani (2015), there is no single rule for the choice of L to cover all applications. Hassani (2007) and Golyandina et al. (2001) notes that in general L should be proportional to the periodicity of the data, large enough to obtain sufficiently separated components but not greater than $N/2$. The output from the embedding step is the trajectory matrix $\mathbf{X}$ which is a Hankel matrix, where all the elements along the diagonal $i + j = const$ are constant (Hassani, 2007):

$$\mathbf{X} = [X_1, \ldots, X_K] = \left(x_{ij} \right)_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{pmatrix}. \tag{2.4}$$

Step 2: Singular Value Decomposition (SVD)

Step two of the decomposition stage is aimed at obtaining the singular values of the trajectory matrix $\mathbf{X}$. These singular values or eigenvalues are able to capture all information in the time series Y_N. In order to obtain the SVD, we need to calculate the matrix $\mathbf{X}\mathbf{X}^T$ which provides us with positive eigenvalues $\lambda_1, \ldots, \lambda_L$

in decreasing order of magnitude. Then, the SVD of $\mathbf{X}$ can be written as:

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_L, \tag{2.5}$$

where $\mathbf{X}_i$ are rank-one bi-orthogonal elementary matrices, $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$, and $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$. Here, U_i and V_i are more commonly known as principal components and represents the left and right eigenvectors of the trajectory matrix $\mathbf{X}$.

The $\sqrt{\lambda_i}$ are also known as the singular values of $\mathbf{X}$ whilst $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_L}\}$ are called the spectrum. The name "Singular Spectrum Analysis" is derived from this property and represents the motive underlying this technique which concentrates on obtaining, and analysing this spectrum of singular values for any given time series in order to identify and differentiate between the signal and noise.

2.2.2 Stage 2: Reconstruction

The reconstruction stage in SSA is concerned with analysing the spectrum of singular values in order to identify and differentiate between the signal and noise, and thereby enable the reconstruction of a less noisy time series which can be used to forecast future data points. The only parameter used at this stage is also the second and final SSA parameter, the number of eigenvalues, r.

Step 1: Grouping

Grouping is the first step in the reconstruction stage. In brief, the grouping step corresponds to splitting the elementary matrices $\mathbf{X}_i$ into several groups and summing the matrices within each group. As noted in Silva and Hassani (2015), if we denote $I = \{i_1, \ldots, i_p\}$ as a group of indices $i_1, \ldots, i_p$, then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. The spilt of

the set of indices $\{1,\ldots,L\}$ into disjoint subsets $I_1,\ldots,I_m$ corresponds to the representation $\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$. The procedure of choosing the sets $I_1,\ldots,I_m$ is called the grouping. For a given group I, the contribution of the component $\mathbf{X}_I$ is measured by the share of the corresponding eigenvalues: $\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$. If the original series contains signal and noise, one then considers two groups of indices, $I_1 = \{1,\ldots,r\}$ and $I_2 = \{r+1,\ldots,L\}$ and associate the group $I = I_1$ with the signal component and the group I_2 with noise.

Note that at the grouping step we have several options for analyzing and differentiating between the signal and noise in a given time series. These include the option of analyzing the periodogram, scatterplot of right eigenvectors or the eigenvalue functions graph (see, Hassani (2007) or Sanei and Hassani (2015)). Once the selection of eigenvalues corresponding to signal and noise is made, it is possible to evaluate the effectiveness of the separation we propose via a statistic known as the weighted correlation (w-correlation). As noted in Golyandina et al. (2001), the w-correlation statistic shows the dependence between two time series and 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, $\parallel Y_N^{(i)} \parallel_w = \sqrt{\left(Y_N^{(i)}, Y_N^{(i)}\right)_w}$, $\left(Y_N^{(i)}, Y_N^{(j)}\right)_w =$

$\sum_{k=1}^{N} w_k y_k^{(i)} y_k^{(j)}$ $(i,j = 1,2)$, $w_k = \min\{k,L,N-k\}$ (here, assume $L \leq N/2$).

Accordingly, if the w-correlation between two reconstructed components are close to 0, this confirms that the corresponding time series are w-orthogonal and that the two components are well separable (Hassani et al., 2009). In contrast, if the w-correlation between two reconstructed components are large, this shows

that the components should be considered as one group.

Step 2: Diagonal averaging

Diagonal averaging is a process which enables one to transform a matrix into a Hankel matrix which can subsequently be converted to a time series, and this is the purpose of the final step in SSA. Sanei and Hassani (2015) elaborates the process concisely as follows. Suppose z_{ij} stands for an element of a matrix $\mathbf{Z}$. Then, the k-th term of the resulting series is obtained by averaging z_{ij} over all i, j such that $i + j = k + 1$. Following diagonal averaging of all matrix components of $\mathbf{X}_{I_j}$ in the expansion of $\mathbf{X}$ above, we end up with another expansion: $\mathbf{X} = \widetilde{\mathbf{X}}_{I_1} + \ldots + \widetilde{\mathbf{X}}_{I_m}$, where $\widetilde{\mathbf{X}}_{I_j}$ is the diagonalized version of the matrix $\mathbf{X}_{I_j}$.

Note that the SVD of the trajectory matrix $\mathbf{X}$ can be represented as:

$$\mathbf{X} = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^T = \mathbf{X}_1 + \ldots, \mathbf{X}_d = \sum_{i \in I} \mathbf{X}_i + \sum_{i \notin I} \mathbf{X}_i,$$

where $d = \max\{i; i = 1, \ldots, L | \lambda_i > 0\}$ (rank $\mathbf{X} = d$), $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ $(i = 1, \ldots, d)$, $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ and $I \subset \{1, \ldots, d\}$. The noise reduced series is reconstructed by $\mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i$ by selecting a set of indices I. However, $\mathbf{X}_I$ does not have a Hankel structure and is not the trajectory matrix of some time series. By performing diagonal averaging over the diagonals $i + j = const$ which corresponds to averaging the matrix elements over the 'antidiagonals' $i + j = k + 1$, the aforementioned issue is overcome: the choice $k = 1$ gives $y_1 = y_{1,1}$, for $k = 2$, $y_2 = (y_{1,2} + y_{2,1})/2$, $y_3 = (y_{1,3} + y_{2,2} + y_{3,1})/3$ and so on. Applying diagonal averaging to the matrix $\mathbf{X}_I$ provides a reconstructed signal s_t, and yields the SSA decomposition of the original series y_t as follows $y_t = s_t + \varepsilon_t$ $(t = 1, 2, \ldots, N)$, where ε_t is the residual series following signal extraction.

2.2.3   Forecasting with SSA

Following the introduction of the basic SSA process which enables filtering and signal extraction in time series, here the two different forecasting approaches in SSA are presented. These are known as Recurrent SSA (SSA-R) and Vector SSA (SSA-V). As a forecasting technique, SSA gives us the option of forecasting either the individual components of the series (which may relate to seasonality or trend for example) or the entire reconstructed series $\hat{Y}_N$ (Hassani and Thomakos, 2010).

According to Sanei and Hassani (2015) the SSA technique can be applied in forecasting any time series that approximately satisfies the linear recurrent formula (LRF):

$$y_j = \sum_{i=1}^{L-1} \alpha_i y_{j-i}, \qquad L \leq j \leq N \tag{2.6}$$

where the coefficients $\alpha_1, \ldots, \alpha_d$ are achieved based on U_i.

The SSA-R forecasting algorithm can be presented as in Golyandina et al. (2001) and Sanei and Hassani (2015).

1. We begin with a time series $Y_N = (y_1, \ldots, y_N)$ of length N.

2. Set L.

3. Consider the linear space $\mathfrak{L}_r \subset \mathbf{R}^L$ of dimension $r < L$. Here, assume that $e_L \notin \mathfrak{L}_r$, where $e_L = (0, 0, \ldots, 1) \in \mathbf{R}^L$.

4. Construct the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K]$ of Y_N.

5. Construct the vectors U_i $(i = 1, \ldots, r)$ from the SVD of $\mathbf{X}$.

6. Compute matrix $\widehat{\mathbf{X}} = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} U_i U_i^T \mathbf{X}$. The vector $\widehat{X}_i$ is the orthogonal projection of X_i onto the space $\mathfrak{L}_r$.

7. Construct the matrix $\widetilde{\mathbf{X}} = \mathscr{H}\widehat{\mathbf{X}} = [\widetilde{X}_1 : \ldots : \widetilde{X}_K]$.

8. Set $v^2 = \pi_1^2 + \ldots + \pi_r^2$, where π_i is the last component of the vector U_i $(i = 1, \ldots, r)$. Moreover, assume that $e_L \notin \mathfrak{L}_r$. This implies that $\mathfrak{L}_r$ is not a vertical space. Therefore, $v^2 < 1$.

9. Determine vector $A = (\alpha_1, \ldots, \alpha_{L-1})$:

$$A = \frac{1}{1 - v^2} \sum_{i=1}^{r} \pi_i U_i^{\nabla},$$

where $U^{\nabla} \in \mathbf{R}^{L-1}$ is the vector consisting of the first $L-1$ components of the vector $U \in \mathbf{R}^L$.

10. Define the time series $Y_{N+h} = (y_1, \ldots, y_{N+h})$ by the formula

$$y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, N \\ \sum_{j=1}^{L-1} \alpha_j y_{i-j} & \text{for } i = N+1, \ldots, N+h \end{cases} \tag{2.7}$$

where $\widetilde{y}_i$ $(i = 1, \ldots, N)$ are the reconstructed series. Then, $y_{N+1}, \ldots, y_{N+h}$ are the h-step ahead recurrent forecasts.

An alternative approach to forecasting with SSA is the SSA-V forecasting algorithm. The main difference between the two approaches is that in SSA-R we consider only the last component of the reconstructed vector for forecasting whereas with SSA-V the entire eigenvector is considered for computing the forecast. The SSA-V approach can be presented as follows, and in doing so Sanei and Hassani (2015) is followed. Consider the following matrix:

$$\Pi = \mathbf{V}^{\nabla}(\mathbf{V}^{\nabla})^T + (1 - v^2)AA^T, \tag{2.8}$$

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

$$\theta^{(v)} : \mathfrak{L}_r \mapsto \mathbf{R}^L, \tag{2.9}$$

where

$$\theta^{(v)}U = \begin{pmatrix} \Pi U^{\triangledown} \\ A^T U^{\triangledown} \end{pmatrix}. \tag{2.10}$$

Define vector Z_i as follows:

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1,\ldots,K \\ \theta^{(v)}Z_{i-1} & \text{for } i = K+1,\ldots,K+h+L-1 \end{cases} \tag{2.11}$$

where, $\widetilde{X}_i$'s are the reconstructed columns of the trajectory matrix after grouping and filtering the noise components. Finally, by constructing matrix $\mathbf{Z} = [Z_1,...,Z_{K+h+L-1}]$ and performing diagonal averaging we can obtain a new series $y_1,...,y_{N+h+L-1}$, where $y_{N+1},...,y_{N+h}$ forms the h-step ahead vector forecasts.

2.3 Multivariate SSA

Where the SSA technique is applied jointly to several series it is referred to as MSSA (Hassani and Mahmoudvand, 2013). According to Sanei and Hassani (2015), the main difference between the recurrent and vector approaches in MSSA is a result of organizing the single trajectory matrix $\mathbf{X}$ of each series into the block trajectory matrix. As such the trajectory matrices can be organized either in vertical or horizontal form. This leads to two different forms of MSSA which are referred to as VMSSA where the vertical form is used and HMSSA where the horizontal form is adopted. Accordingly, there are four different MSSA forecasting algorithms as shown below (Hassani and Mahmoudvand, 2013).

$$\text{MSSA forecasting approach} = \begin{cases} \text{HMSSA} & \begin{cases} \text{Recurrent approach} \\ \\ \text{Vector approach} \end{cases} \\ \\ \text{VMSSA} & \begin{cases} \text{Recurrent approach} \\ \\ \text{Vector approach} \end{cases} \end{cases}$$

In what follows, the theory underlying VMSSA and HMSSA are presented by following the representations in Hassani and Mahmoudvand (2013) and Sanei and Hassani (2015).

2.3.1 Vertical MSSA (VMSSA)

Consider M time series with different series length N_i; $Y_{N_i}^{(i)} = (y_1^{(i)}, \ldots, y_{N_i}^{(i)})$ $(i = 1, \ldots, M)$. Note that the univariate form can be acquired by setting $M = 1$ for all multivariate algorithms considered in this chapter.

Stage 1: Decomposition

Step 1: Embedding.

Embedding, as previously mentioned is a mapping that transfers a one-dimensional time series $Y_{N_i}^{(i)} = (y_1^{(i)}, \ldots, y_{N_i}^{(i)})$ into a multidimensional matrix $[X_1^{(i)}, \ldots, X_{K_i}^{(i)}]$ with vectors $X_j^{(i)} = (y_j^{(i)}, \ldots, y_{j+L_i+1}^{(i)})^T \in \mathbf{R}^{L_i}$, where L_i $(2 \leq L_i \leq N_i - 1)$ is the window length for each series with length N_i and $K_i = N_i - L_i + 1$. The output from the embedding step is the trajectory matrix (which is a Hankel matrix) $\mathbf{X}^{(i)} = [X_1^{(i)}, \ldots, X_{K_i}^{(i)}] = (x_{mn})_{m,n=1}^{L_i,K_i}$. Therefore applying the above procedure for each series separately provides M different $L_i \times K_i$ trajectory matrices $\mathbf{X}^{(i)}$

$(i = 1, \ldots, M)$. In order to form a block Hankel matrix in a vertical form, it is required to have $K_1 = \ldots, K_M = K$. Note that VMSSA enables us to have various window length L_i and different series length N_i, but similar K_i for all series. The result of this step is the following block Hankel trajectory matrix

$$\mathbf{X}_V = \begin{bmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(M)} \end{bmatrix},$$

where $\mathbf{X}_V$, the output of the first step is a block Hankel trajectory matrix formed in *vertical* form.

Step 2: Singular Value Decomposition (SVD)

This step performs the SVD of $\mathbf{X}_V$. If we denote $\lambda_{V_1}, \ldots, \lambda_{V_{L_{sum}}}$ as the eigenvalues of $\mathbf{X}_V \mathbf{X}_V^T$, arranged in decreasing order $\left(\lambda_{V_1} \geq \ldots \lambda_{V_{L_{sum}}} \geq 0 \right)$ and $U_{V_1}, \ldots, U_{V_{L_{sum}}}$, the corresponding eigenvectors, where $L_{sum} = \sum_{i=1}^M L_i$, then the structure of the matrix $\mathbf{X}_V \mathbf{X}_V^T$ is as follows:

$$\mathbf{X}_V \mathbf{X}_V^T = \begin{bmatrix} \mathbf{X}^{(1)} \mathbf{X}^{(1)T} & \mathbf{X}^{(1)} \mathbf{X}^{(2)T} & \cdots & \mathbf{X}^{(1)} \mathbf{X}^{(M)T} \\ \mathbf{X}^{(2)} \mathbf{X}^{(1)T} & \mathbf{X}^{(2)} \mathbf{X}^{(2)T} & \cdots & \mathbf{X}^{(2)} \mathbf{X}^{(M)T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}^{(M)} \mathbf{X}^{(1)T} & \mathbf{X}^{(M)} \mathbf{X}^{(2)T} & \cdots & \mathbf{X}^{(M)} \mathbf{X}^{(M)T} \end{bmatrix}. \tag{2.12}$$

Note that the matrix $\mathbf{X}^{(i)} \mathbf{X}^{(i)T}$, which is used in univariate SSA, for the series $Y_{N_i}^{(i)}$, appears along the main diagonal and the products of two Hankel matrices $\mathbf{X}^{(i)} \mathbf{X}^{(j)T} (i \neq j)$, which are related to the series $Y_{N_i}^{(i)}$ and $Y_{N_j}^{(j)}$, appears in the off-diagonal. The SVD of $\mathbf{X}_V$ can be written as $\mathbf{X}_V = \mathbf{X}_{V_1} + \cdots + \mathbf{X}_{V_{L_{sum}}}$, where $\mathbf{X}_{V_i} = \sqrt{\lambda_i} U_{V_i} V_{V_i}^T$ and $V_{V_i} = \mathbf{X}_V^T U_{V_i} / \sqrt{\lambda_{V_i}}$ ($\mathbf{X}_{V_i} = 0$ if $\lambda_{V_i} = 0$).

Stage 2: Reconstruction

Step 1: Grouping

The grouping step, as with univariate SSA, corresponds to splitting the matrices $\mathbf{X}_{V_1}, \ldots, \mathbf{X}_{V_{L_{sum}}}$ into several disjoint groups and summing the matrices within each group. The split of the set of indices $\{1, \ldots, L_{sum}\}$ into disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation $\mathbf{X}_v = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$. The procedure of choosing the sets $I_1, \ldots, I_m$ is termed grouping. Assuming that we have only signal and noise components, we use two groups of indices, $I_1 = \{1, \ldots, r\}$ and $I_2 = \{r+1, \ldots, L_{sum}\}$ such that the group $I = I_1$ is associated with signal component and the group I_2 with noise.

Step 2: Diagonal averaging or Hankelization.

Diagonal averaging is used to transform the reconstructed matrix $\widehat{\mathbf{X}}_{V_i}$ to the form of a Hankel matrix, which can be subsequently converted to a time series. Let $\widetilde{\mathbf{X}}^{(i)}$ be the approximation of $\mathbf{X}^{(i)}$ obtained following diagonal averaging. If $\widetilde{x}_{mn}^{(i)}$ stands for an element of a matrix $\widetilde{\mathbf{X}}^{(i)}$, then the j-th term of the reconstructed series $\widetilde{Y}_{N_i}^{(i)} = (\widetilde{y}_1^{(i)}, \ldots, \widetilde{y}_j^{(i)}, \ldots, \widetilde{y}_{N_i}^{(i)})$ is achieved by arithmetic averaging $\widetilde{x}_{mn}^{(i)}$ over all (m, n) such that $m + n - 1 = j$.

2.3.2 Horizontal MSSA (HMSSA)

The decomposition and reconstruction stages of the HMSSA algorithm are similar to those provided above for VMSSA except for the structure of the block Hankel matrix. Assume that we have M different $L_i \times K_i$ trajectory matrices $\mathbf{X}^{(i)}$ ($i = 1, \ldots, M$). To construct a block Hankel matrix in the horizontal form we need to have $L_1 = L_2 = \ldots = L_M = L$. This means that we have different values of K_i and series length N_i, but similar L_i. The result of this step is as follows:

$$\mathbf{X}_H = \left[\ \mathbf{X}^{(1)} : \ \ \mathbf{X}^{(2)} : \ \ \cdots \ \ : \mathbf{X}^{(M)} \ \right].$$

Hence, the structure of the matrix $\mathbf{X}_H\mathbf{X}_H^T$ is such that:

$$\mathbf{X}_H\mathbf{X}_H^T = \mathbf{X}^{(1)}\mathbf{X}^{(1)^T} + \cdots + \mathbf{X}^{(M)}\mathbf{X}^{(M)^T}. \tag{2.13}$$

The structure of the matrix $\mathbf{X}_H\mathbf{X}_H^T$ implies that in HMSSA, we do not have any cross-product between Hankel matrices $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$. Moreover, in this format, the sum of $\mathbf{X}^{(i)}\mathbf{X}^{(i)^T}$ provides the block Hankel matrix. Note also that performing the SVD of $\mathbf{X}_H$ in HMSSA yields L eigenvalues as with SSA, whilst we have $L_{sum} = \sum_{i=1}^{M} L_i$ eigenvalues in VMSSA.

2.3.3 Forecasting with MSSA

VMSSA Recurrent (VMSSA-R) Forecasting Algorithm

Let us have M series $Y_{N_i}^{(i)} = (y_1^{(i)}, \ldots, y_{N_i}^{(i)})$ and corresponding window length L_i, $1 < L_i < N_i, i = 1 \ldots, M$. Then, the h-step ahead VMSSA-R forecasting algorithm is as follows (Hassani and Mahmoudvand, 2013; Sanei and Hassani, 2015).

1. For a fixed value of K, construct the trajectory matrix $\mathbf{X}^{(i)} = [X_1^{(i)}, \ldots, X_K^{(i)}] = (x_{mn})_{m,n=1}^{L_i,K}$ for each single series $Y_{N_i}^{(i)}$ $(i = 1, \ldots, M)$ separately.

2. Construct the block trajectory matrix $\mathbf{X}_V$ as follows:

$$\mathbf{X}_V = \begin{bmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(M)} \end{bmatrix}.$$

3. Let $\mathbf{U}_{V_j} = (U_j^{(1)}, \ldots, U_j^{(M)})^T$ be the jth eigenvector of the $\mathbf{X}_V \mathbf{X}_V^T$, where $U_j^{(i)}$ with length L_i corresponds to the series $Y_{N_i}^{(i)}$ $(i = 1, \ldots, M)$.

4. Consider $\widehat{\mathbf{X}}_V = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} U_{V_i} U_{V_i}^T \mathbf{X}_V$ as the reconstructed matrix achieved from r eigentriples:

$$\widehat{\mathbf{X}}_V = \begin{bmatrix} \widehat{\mathbf{X}}^{(1)} \\ \vdots \\ \widehat{\mathbf{X}}^{(M)} \end{bmatrix}.$$

5. Consider matrix $\widetilde{\mathbf{X}}^{(i)} = \mathscr{H} \widehat{\mathbf{X}}^{(i)}$ $(i = 1, \ldots, M)$ as the result of the Hankelization procedure of the matrix $\widehat{\mathbf{X}}^{(i)}$ obtained from the previous step, where $\mathscr{H}$ is a Hankel operator.

6. Assume $U_j^{(i)\nabla}$ denotes the vector of the first $L_i - 1$ components of the vector $U_j^{(i)}$ and $\pi_j^{(i)}$ is the last component of the vector $U_j^{(i)}$ $(i = 1, \ldots, M)$.

7. Select the number of r eigentriples for the reconstruction stage that can also be used for forecasting purposes.

8. Define matrix $\mathbf{U}^{\nabla M} = (U_1^{\nabla M}, \ldots, U_r^{\nabla M})$, where $U_j^{\nabla M}$ is as follows:

$$U_j^{\nabla M} = \begin{bmatrix} U_j^{(1)\nabla} \\ \vdots \\ U_j^{(M)\nabla} \end{bmatrix}.$$

9. Define matrix $\mathbf{W}$ as follows:

$$\mathbf{W} = \begin{bmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \cdots & \pi_r^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \cdots & \pi_r^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ \pi_1^{(M)} & \pi_2^{(M)} & \cdots & \pi_r^{(M)} \end{bmatrix}.$$

10. If the matrix $\left(\mathbf{I}_{M \times M} - \mathbf{W}\mathbf{W}^T\right)^{-1}$ exists and $r \leq L_{sum} - M$, then the h-step ahead VMSSA forecasts exist and is achieved by the following formula:

$$\begin{cases} \left[\tilde{y}_{j_1}^{(1)}, \ldots, \tilde{y}_{j_M}^{(M)}\right], & j_i = 1, \ldots, N_i \\ \\ \left(\mathbf{I}_{M \times M} - \mathbf{W}\mathbf{W}^T\right)^{-1} \mathbf{W}\mathbf{U}^{\nabla M^T}\mathbf{Z}_h, & j_i = N_i + 1, \ldots, N_i + h, \end{cases} \tag{2.14}$$

where, $\mathbf{Z}_h = \left[Z_h^{(1)}, \ldots, Z_h^{(M)}\right]^T$ and $Z_h^{(i)} = \left[\hat{y}_{N_i-L_i+h+1}^{(i)}, \ldots, \hat{y}_{N_i+h-1}^{(i)}\right]$ $(i = 1, \ldots, M)$. It should be noted that equation (4.10) indicates that the h-step ahead forecasts of the refined series $\hat{Y}_{N_i}^{(i)}$ are obtained by a multi dimensional linear recurrent formula (LRF). For the univariate case, there is only the one dimensional LRF.

HMSSA Recurrent (HMSSA-R) Forecasting Algorithm

1. For a fixed value of L, construct the trajectory matrix $\mathbf{X}^{(i)} = [X_1^{(i)}, \ldots, X_K^{(i)}] = (x_{mn})_{m,n=1}^{L,K_i}$ for each single series $Y_{N_i}^{(i)}$ $(i = 1, \ldots, M)$ separately.

2. Construct the block trajectory matrix $\mathbf{X}_H$ as follows:

$$\mathbf{X}_H = \begin{bmatrix} \mathbf{X}^{(1)} : & \mathbf{X}^{(2)} : & \cdots & : \mathbf{X}^{(M)} \end{bmatrix}.$$

3. Let vector $U_{H_j} = (u_{1j}, \ldots, u_{Lj})^T$, with length L, be the jth eigenvector of $\mathbf{X}_H\mathbf{X}_H^T$.

4. Consider $\widehat{\mathbf{X}}_H = \sum_{i=1}^{r} U_{H_i} U_{H_i}^T \mathbf{X}_H$ as the reconstructed matrix obtained using r eigentriples:

$$\mathbf{X}_H = \left[\begin{array}{cccc} \widehat{\mathbf{X}}^{(1)} : & \widehat{\mathbf{X}}^{(2)} : & \cdots & : \widehat{\mathbf{X}}^{(M)} \end{array} \right].$$

5. Consider matrix $\widetilde{\mathbf{X}}^{(i)} = \mathscr{H}\widehat{\mathbf{X}}^{(i)}$ ($i = 1, \ldots, M$) as the result of the Hankelization procedure of the matrix $\widehat{\mathbf{X}}^{(i)}$ obtained from the previous step.

6. Let $U_{H_j}^{\triangledown}$ denotes the vector of the first $L-1$ coordinates of the eigenvectors U_{H_j}, and π_{H_j} indicates the last coordinate of the eigenvectors U_{H_j} ($j = 1, \ldots, r$).

7. Define $\upsilon^2 = \sum_{j=1}^{r} \pi_{H_j}^2$.

8. Denote the linear coefficients vector $\mathscr{R}$ as follows:

$$\mathscr{R} = \frac{1}{1 - \upsilon^2} \sum_{j=1}^{r} \pi_{Hj} U_{Hj}^{\triangledown}. \tag{2.15}$$

9. If $\upsilon^2 < 1$, then the h-step ahead HMSSA forecasts exist and can be calculated by the formula:

$$\left[\hat{y}_{j_1}^{(1)}, \ldots, \hat{y}_{j_M}^{(M)} \right]^T = \begin{cases} \left[\tilde{y}_{j_1}^{(1)}, \ldots, \tilde{y}_{j_M}^{(M)} \right], & j_i = 1, \ldots, N_i, \\ \\ \mathscr{R}^T \mathbf{Z}_h, & j_i = N_i + 1, \ldots, N_i + h, \end{cases} \tag{2.16}$$

where, $\mathbf{Z}_h = \left[Z_h^{(1)}, \ldots, Z_h^{(M)} \right]^T$ and $Z_h^{(i)} = \left[\hat{y}_{N_i-L+h+1}^{(i)}, \ldots, \hat{y}_{N_i+h-1}^{(i)} \right]$ ($i = 1, \ldots, M$).

Note that equation (5.8) indicates that the *h*-step ahead forecasts of each series are achieved by the same LRF generated considering all series in a multivariate system. In what follows, the MSSA Vector forecasting algorithms are explained by following Hassani and Mahmoudvand (2013), the authors who introduced these two algorithms.

HMSSA Vector (HMSSA-V) Forecasting Algorithm

The procedure for HMSSA-V is very similar to its univariate version, SSA-V and HMSSA-R. We begin by following items (1)-(7) of HMSSA-R. Then, consider the following matrix

$$\Pi = \mathbf{U}^\nabla \mathbf{U}^{\nabla T} + (1 - v^2) RR^T, \tag{2.17}$$

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

$$\mathscr{P}^{(v)} : \mathfrak{L}_r \mapsto \mathbb{R}^L, \tag{2.18}$$

where

$$\mathscr{P}^{(v)} Y = \begin{pmatrix} \Pi Y_\triangle \\ R^T Y_\triangle \end{pmatrix}, Y \in \mathfrak{L}_r, \tag{2.19}$$

and $Y_\triangle$ is vector of last $L - 1$ elements of Y.

1. Define vector $Z_j^{(i)}$ ($i = 1, \dots, M$) as follows:

$$Z_j^{(i)} = \begin{cases} \widetilde{X}_j^{(i)} & \text{for } j = 1, \dots, k_i \\ \mathscr{P}^{(v)} Z_{j-1}^{(i)} & \text{for } j = k_i + 1, \dots, k_i + h + L - 1 \end{cases} \tag{2.20}$$

where, $\widetilde{X}_j^{(i)}$'s are the reconstructed columns of trajectory matrix of the ith series after grouping and leaving noise components.

2. Now, by constructing matrix $\mathbf{Z}^{(i)} = [Z_1^{(i)}, ..., Z_{k_i+h+L-1}^{(i)}]$ and performing diagonal averaging we obtain a new series $\hat{y}_1^{(i)}, ..., \hat{y}_{N_i+h+L-1}^{(i)}$, where $\hat{y}_{N_i+1}^{(i)}, ..., \hat{y}_{N_i+h}^{(i)}$ provides the h-step ahead of HMSSA-V forecast.

VMSSA Vector (VMSSA-V) Forecasting Algorithm

Begin by considering items (1)-(10) of VMSSA-R. Consider the matrix:

$$\Pi = \mathbf{U}^{\triangledown}\mathbf{U}^{\triangledown T} + \mathscr{R}\left(\mathbf{I}_{M \times M} - \mathbf{W}\mathbf{W}^T\right)\mathscr{R}^T, \qquad (2.21)$$

where, $\mathscr{R} = \mathbf{U}^{\triangledown}\mathbf{W}^T\left(\mathbf{I}_{M \times M} - \mathbf{W}\mathbf{W}^T\right)^{-1}$. The following algorithm is proposed for calculating the VMSSA-V forecasts (see, Hassani and Mahmoudvand (2013) for theorem and proof).

1. Define vectors Z_i as follows:

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, k \\ \mathscr{P}^{(v)}Z_{i-1} & \text{for } i = k+1, \ldots, k+h+L_{\max}-1, \end{cases} \qquad (2.22)$$

where, $L_{\max} = \max\{L_1, \ldots, L_M\}$.

2. Constructing the matrix $\mathbf{Z} = [Z_1 : \ldots : Z_{K+h+L_{\max}-1}]$ and making its hankelization. Using this calculation we obtain $\hat{y}_1^{(i)}, \ldots, \hat{y}_{N+h+L_{\max}}^{(i)}$ $(i = 1, \ldots, M)$.

3. The numbers $\hat{y}_{N_i+1}^{(i)}, \ldots, \hat{y}_{N_i+h}^{(i)}$ $(i = 1, \ldots, M)$ form the h step ahead VMSSA-V forecasts.

Given that there are two different MSSA approaches it is pertinent to note their similarities and differences which can be useful when choosing between

them. Hassani and Mahmoudvand (2013)considers various perspectives such as series length, the value of the Window Length (L_i), the number of nonzero singular values obtained from the block trajectory matrix and LRF. Table 2.1 presents this summary. As evident from this table there are some restrictions on selecting values of L and K depending on the MSSA approach used. However, to this date there is no definitive study that notes which MSSA approach is best. In terms of selecting between the Recurrent or Vector forecasting approaches, the Vector forecasting algorithm is known to be more robust than the Recurrent forecasting algorithm if there are outliers in the series being analysed (Hassani et al., 2014).

Table 2.1 Similarities and dissimilarities between the VMSSA and HMSSA algorithms.

Method	Series Length	L_i	K_i	Number of λ_i	LRF
VMSSA	Different	Different	Equal	$\sum L_i$	Different
HMSSA	Different	Equal	Different	L	Equal

2.4 Benchmark Forecasting Models

As explained in Chapter 1, automated forecasting models are becoming increasingly popular in the modern age. Given that this thesis seeks to automate and optimize the SSA and MSSA techniques, selected as benchmark models for comparison purposes are the automated forecasting algorithms for ARIMA, Holt-Winters, ETS and Neural Networks as provided via the forecast package in `R`. In addition, the choice of these benchmark models have also been influenced by previous applications in literature. However, it is important to note that the applications which follow do not intend on presenting the newly proposed SSA and MSSA approaches as universally best at this time. In terms of the forecasting strategy, unless stated otherwise the applications which follow exploit

a recursive forecasting strategy. Details on the selected benchmark models are concisely presented below and in doing so Ghodsi et al. (2015) is mainly followed.

2.4.1   Autoregressive Integrated Moving Average (ARIMA)

ARIMA is recognized as one of the most popular benchmark forecasting techniques. Used in this research is *auto.arima* which is an optimized version of the ARIMA model and provided via the forecast package in R. A detailed description of the algorithm can be found in Hyndman and Khandakar (2008). In brief, the number of differences d is determined using either a KPSS test, Augmented Dickey Fuller test or the Phillips-Perron test. Thereafter, the algorithm minimises the Akaike Information Criterion (AIC) to determine the values for the order of autoregressive terms p, and the order of the moving average process q. The optimal model is chosen to be the model which represents the smallest AIC. The decision on the inclusion or exclusion of the constant c is dependent on the value of d.

According to Hyndman and Athanasopoulos (2012) a non-seasonal ARIMA model may be written as:

$$(1 - \phi_1 B - \ldots \phi_p B^p)(1-B)^d y_t = c + (1 + \phi_1 B + \ldots + \phi_q B^q)e_t, \qquad (2.23)$$

or

$$(1 - \phi_1 B - \ldots \phi_p B^p)(1-B)^d (y_t - \mu t^d / d!) = (1 + \phi_1 B + \ldots + \phi_q B^q)e_t, \quad (2.24)$$

where μ is the mean of $(1-B)^d(y_t$, $c = \mu(1 - \phi_1 - \ldots - \phi_p)$ and B is the back-

shift operator. In R, the inclusion of a constant in a non-stationary ARIMA model is equivalent to inducing a polynomial trend of order d in the forecast function and when $d=0$, μ is the mean of y_t. Likewise, Hyndman and Khandakar (2008) presents the seasonal ARIMA model as:

$$\Phi(B^m)\phi(B)(1-B^m)^D(1-B)^d y_t = c + \Theta(B^m)\theta(B)\varepsilon_t, \qquad (2.25)$$

where $\Phi(z)$ and $\Theta(z)$ are the polynomials of orders P and Q, and ε_t is white noise. Note that if $c \neq 0$, there is an implied polynomial of order $d+D$ in the forecast function. In order to determine the values of p and q the AIC of the following form is minimised:

$$AIC = -2log(l) + 2(p+q+P+Q+k), \qquad (2.26)$$

where $k = 1$ if $c \neq 0$ and 0 otherwise, and l represents the maximum likelihood of the fitted model.

2.4.2   Holt-Winters (HW)

The Holt-Winters models is a popular time series analysis and forecasting technique which continues to be used by Central Banks around the globe. It was developed through the work by Holt in 1957 as published in Holt (2004) and Winters (1960). The R software allows for calculating forecasts from the HW model via the stats package.

The HW forecasting equations are presented below, and in doing so Holt (2004) and Winters (1960) are followed. The additive HW prediction function (for a time series with period length p) is

$$\hat{Y}_{t+h} = a_t + h * b_t + s_{[t-p+1+(h-1)mod\,p]}, \qquad (2.27)$$

where a_t, b_t and s_t are given by

$$a_t = \alpha(Y_t - s_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1}), \qquad (2.28)$$

$$b_t = \beta(a_t - a_{t-1}) + (a - \beta)b_{t-1}, \qquad (2.29)$$

$$s_t = \gamma(Y_t - a_t) + (1 - \gamma)s_{t-p}. \qquad (2.30)$$

The multiplicative HW prediction function (for a time series with period length p) is

$$\hat{Y}_{t+h} = (a_t + h * b_t) * s_{[t-p+1+(h-1)mod\,p]}, \qquad (2.31)$$

where a_t, b_t and s_t are given by

$$a_t = \alpha(Y_t/s_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1}), \qquad (2.32)$$

$$b_t = \beta(a_t - a_{t-1}) + (a - \beta)b_{t-1}, \qquad (2.33)$$

$$s_t = \gamma(Y_t/a_t) + (1 - \gamma)s_{t-p}. \qquad (2.34)$$

The algorithm is programmed to find the optimal values of α, β and γ by minimizing the squared one-step prediction error [1].

---

[1] Those interested in the details of the algorithm are referred to https://stat.ethz.ch/R-manual/R-devel/library/stats/html/HoltWinters.html

2.4.3   Exponential Smoothing (ETS)

The ETS technique in the forecast package in R overcomes a limitation found in earlier exponential smoothing models which failed to provide a method for easy calculation of prediction intervals (Makridakis et al., 1998). A detailed description of ETS can be found in (Hyndman and Athanasopoulos, 2012). In brief, this ETS model takes into account the error, trend and seasonal components along with over 30 possible options for choosing the best exponential smoothing model via optimization of initial values and parameters using the MLE, and selecting the best model based on the AIC. Figure 2.2 summarises the several ETS formulae that are evaluated in the forecast package to select the best model to fit the data. Note that in this figure, ell_t denotes the series level at time t, b_t denotes the slope, s_t denotes the seasonal component of the series, and m denotes the number of seasons in a year; α, β, γ and ϕ are smoothing parameters, $\phi_h = \phi + \phi_2 + \ldots + \phi^h$ and $h_m^+ = [(h-1) mod m] + 1$ (Hyndman and Athanasopoulos, 2012).

ADDITIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A$_d$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M$_d$	$y_t = \ell_{t-1}b_{t-1}^{\phi} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^{\phi} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^{\phi})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A$_d$	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M$_d$	$y_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Fig. 2.2 State space equations for each of the models in the ETS framework (Hyndman and Athanasopoulos, 2012).

2.4.4 Neural Networks (NN)

The NN model in the forecast package in R is referred to as *nnetar*. A detailed description of the model can be found in Hyndman and Athanasopoulos (2012) along with an explanation of the underlying dynamics. In brief, the *nnetar* function trains 25 neural networks by adopting random starting values and then obtains the mean of the resulting predictions to compute the forecasts. The neural network takes the form

$$\hat{y}_t = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j \psi(x_t^T . \hat{\gamma}_j), \tag{2.35}$$

where x_t consist of p lags of y_t and ^{T} denotes transpose. Then, the function ψ has the logistic form

$$\psi(x_t^{'} . \hat{\gamma}_j) = [1 + exp(-\hat{\gamma}_{j0} + \sum_{i=1}^{p} \hat{\gamma}_{ji} . y_{t-1})]^{-1} . j = 1, \ldots, k \tag{2.36}$$

This form of neural networks is referred to as a one hidden layer feed forward neural network model. The nonlinearity arises through the lagged y_t entering in a flexible way through the logistic functions of (2.28). The number of logistic functions (k) included, is known as the number of hidden nodes. The parameters in the neural network model are selected based on a loss function embedded into learning algorithm. It may be noted that in all cases the selected neural network model has only k=1 hidden node, p=2 lags.

2.5 Metrics

Presented in this section are the various metrics which are used to compare the forecasting results obtained via the many applications which follow. This thesis considers both loss functions and direction of change criterions for comparing

between forecasts. This is because a loss function should be coupled with a criterion such as direction of change in order to determine if a forecast is reliable enough for decision making. Also it is possible that those interested in identifying business cycle changes such as recessions or expansions would prefer a criterion such as direction of change to be more important and useful than a loss function alone.

2.5.1   Root Mean Squared Error (RMSE)

The applications which follow rely mainly on the RMSE as a metric. The choice of RMSE as the main criterion is for several reasons. Firstly, the RMSE continues to remain a popular measure of forecast accuracy (see, for example, Zhang et al. (1998), Hassani et al. (2009), and Hassani et al. (2013b)). Secondly, the RMSE is able to indicate the error in the same units as the original data. Given that each application which follows considers comparing between data sets with identical units it is easier to compare between the forecasts by relying on the RMSE which is also easier to interpret in relation to business decisions (Armstrong and Collopy, 1992). Thirdly, the applications which follow requires comparisons between forecast errors with a Gaussian distribution, and Chai and Draxler (2014) notes that the RMSE is better at representing model performance when the error distribution is expected to be Gaussian. As an example, the RMSE ratios of SSA to that of ETS are provided:

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

where, $\widehat{y}_{T+h}$ represents the h-step ahead forecast obtained by SSA, $\widetilde{y}_{T+h}$ is the h-step ahead forecast from the ETS model, and N is the number of the forecasts.

If $\frac{SSA}{ETS}$ is less than 1, then the SSA outperforms ETS by 1-$\frac{SSA}{ETS}$ percent.

2.5.2   Mean Absolute Percentage Error (MAPE)

The MAPE measure is also used in this thesis for quantifying forecast accuracy. In brief, the lower the MAPE value, the more accurate the forecast.

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} |100 \times \frac{y_{T+h} - \widehat{y}_{T+h,i}}{y_{T+h}}|,$$

where y_{T+h} represents the actual data corresponding to the h step ahead forecast, and $\widehat{y}_{T+h,i}$ is the h step ahead forecasts obtained from a particular forecasting model.

2.5.3   Direction of Change (DC)

The DC criterion is a measure of the percentage of forecasts that accurately predict the direction of change (Hassani et al., 2013b; Hassani and Thomakos, 2010). Here, the concept of DC is explained in brief by following Hassani et al. (2013b).

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. As noted in Hassani and Thomakos (2010), based on the Moivre-Laplace central limit theorem, for large samples, the test statistic $2(\tilde{D}_X - 0.5)n^{1/2}$ is approximately distributed as standard normal. Where the results for the DC criterion are statistically significant, it shows whether they are significantly greater than the pure chance (Hassani et al., 2013a). Accordingly, if $\tilde{D}_X$ is significantly greater than 0.5, then the forecast is said to have the ability of predicting the DC, and if $\tilde{D}_X$ is significantly less than 0.5, then the forecast tends to give an incorrect DC (Hassani and

Thomakos, 2010).

Several authors have discussed the importance of the DC criterion as a measure of forecast accuracy. In particular, Ash et al. (1997) are of the view that a smaller prediction error and a misforecasted direction of change is more problematic than a larger directionally correct error for some purposes. Clements and Smith (1999) subscribe to a similar view as they note that the DC criterion is a better measure of the quality of forecasts. However, Heravi et al. (2004) are more explicit when they state that the DC criterion is particularly important for capturing business cycle fluctuations pertaining to recessions and expansions.

2.5.4 Diebold-Mariano (DM) Test

One of the statistical tests considered as a measure for comparing between the predictive accuracy of two sets of forecasts in this thesis is the DM test. The DM test was introduced by Diebold and Mariano (1995) but was improved through the work of Harvey et al. (1997) whereby the authors sought to overcome several issues with the original DM test. The modified DM test statistic is (Harvey et al., 1997):

$$[\frac{n+1-2h+n^{-1}h(h-1)}{n}]^{1/2} * S_i, \qquad (2.37)$$

where S_i is the original DM statistic which is explained in detail in Chapter 3 and is therefore not reproduced here. The hypothesis of the test used here are:

$$H_0 : E(d_t) = 0, H_1 : E(d_t) \neq 0. \qquad (2.38)$$

where the null hypothesis H_0 states that both forecasts have the same accuracy and the alternative hypothesis H_1 states that the two forecasts have different levels of accuracy. Note that d_t is the loss differential between two different

forecasts.

Chapter 3

A Kolmogorov-Smirnov based Test for Comparing the Predictive Accuracy of Two Sets of Forecasts

Presented in this chapter is the first contribution of this thesis which is a complement statistical test for comparing between the predictive accuracy of two sets of forecasts. This test has been founded upon the principles of cumulative distribution functions and stochastic processes.

3.1 Introduction

There is a consensus that any attempt to justify the comparative superiority of forecasts from a given model is both incomplete and inadmissible if no consideration has been given to the statistical significance associated with the comparison. Tests on forecast evaluation and comparison have a long and detailed history which can be found in Chapter 3 of Elliot and Timmermann (2013). Few historically popular examples of such statistical tests are discussed in Christiano (1989); Diebold and Mariano (1995); Meese and Rogoff (1988) and Harvey

et al. (1997). Of these, the Diebold-Mariano (DM) test (Diebold and Mariano, 1995) is one which is highly cited, and its popularity is evident via statements such as that in Diebold (2013), pp.8 according to which, "for comparing forecasts, DM is the only game in town."

Whilst there is indeed no question regarding the popularity of the DM test, it is pertinent to note that the DM test is by no means a panacea. At present there exists other improved variants for evaluating the statistical significance between forecasts. Two sound examples would be Hansen's Hansen (2005) Superior Predictive Ability (SPA) test, and Hansen et al.'s Hansen et al. (2011) Model Confidence Set (MCS) which are superior to the DM test. In addition, recently there has been a renewed interest in research focussing on testing the predictive accuracy of forecasts through the work of Clark and McCracken (2009, 2012); Gilleland and Roux (2015); Gneiting and Raftery (2007). Clark and McCracken Clark and McCracken (2012) in particular shows that the DM test is inferior or inappropriate for use alongside nested forecasting models.

The aim of this chapter is to introduce a complement statistical test (which differs from the tests noted above) for comparing between the predictive accuracy of forecasts whilst overcoming the constraints of the DM test which are identified below. Interestingly, regardless of the existence of more superior tests, the DM test continues to be cited in forecasting literature both in isolation and at times along side SPA and MCS tests, see for example Hassani et al. (2015); Silva and Hassani (2015). This research uses the DM test as a benchmark with the reasons being justified in what follows.

The DM test can be briefly introduced as an asymptotic *z*-test for the hypothesis that the loss differential is zero (Diebold, 2013)[1]. Whilst it is not the intention of this research to ridicule any proven test currently adopted for com-

---

[1]Note that the Granger and Newbold (1977) assumption of forecast errors having zero mean is not essential according to Morgan (1939).

paring the accuracy of forecasts, it is evident that the need for a complement statistical test arises owing to the following reasons which relate to both theoretical and empirical issues with the DM test. Firstly, the original DM test was limited by finite sample properties (Diebold and Mariano, 1995). Secondly, as a parametric test, the DM test requires that the loss differential has a stationary covariance (Diebold, 2013). The failure to meet this assumption invalidates the results and imposes a restriction on the applicability of this test. These issues were later addressed in Harvey et al. (1997) when a solution was achieved via the inclusion of a new assumption whereby all autocovariances of the mean loss differential beyond some lag length are assumed to be 0. However, according to the recent findings in Hassani (2010) and Hassani et al. (2012a) it has been proven that when the lag of a sample autocorrelation function (ACF) exceeds 1, the sum of the ACF is always equal to $-\frac{1}{2}$. In fact, according to Harvey et al. (1997) the modified DM statistic continues to be multiplied by the original DM statistic $[\hat{V}(\bar{d})]^{-0.5}\bar{d}$, where $\hat{V}(\bar{d}) \approx n^{-1}[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k]$ and γ_k is the kth autocovariance of d_t. Then, as per recent findings (Hassani, 2010; Hassani et al., 2012a) it implies that the sum of the autocovariance, $\sum_{k=1}^{h-1}\gamma_k = -\frac{1}{2}\gamma_0$ which in turn ensures that the expectation of $\hat{V}(\bar{d})=0$, and therefore the modified DM test statistic tends to infinity. Thus, if two models are used to forecast *n* data points without repeating or updating the data, then the modified DM test cannot be applied as the sum of the covariance will be zero. Thirdly, the modified DM test statistic for improved small sample properties is dependent on the Student's *t* distribution (Harvey et al., 1997) which cannot be justified unless the forecast errors are independent and normally distributed. In addition, even though Harvey et al. (1997) asserts that the modified DM test can provide efficient results when faced with small sample properties, in practice there can be instances when this assertion fails to hold. For example, in some instances where the Ratio of the Root Mean Squared Error (RRMSE) criterion shows that the forecasts from a partic-

ular model are for example 60% more accurate than the forecasts from another
model (with a large sample size), the DM test fails to show a statistically signif-
icant difference between such forecasts. Moreover, when faced with comparing
for example a small sample of $h = 12$ steps ahead forecasts there is a tendency
for the modified DM test to always report a significant difference between fore-
casts even when the RRMSE criterion is at around 99%. Finally, according to
the simulation results reported in Harvey et al. (1997) the modified DM test is
not accurately sized for both small and large samples beyond the one-step ahead
forecasting horizon.

The proposed test is founded upon the principles of the Kolmogorov-Smirnov
(KS) test (Kolmogorov, 1933) and is non-parametric in nature. The choice of
a non-parametric test is important as in the real world we are mostly faced
with data which fails to meet the assumptions of normality and stationarity un-
derlying parametric tests. The proposed test (referred to as the Kolmogorov-
Smirnov Predictive Accuracy or KSPA test) was motivated by the work of Has-
sani et al. (2009) and Hassani et al. (2013b), where cumulative distribution func-
tions (c.d.f.'s) relating to the absolute value of forecast errors are exploited to
determine if one forecasting technique provides superior forecasts in compari-
son to another technique. The approach presented in the aforementioned papers
are in fact based on the concept of stochastic dominance. However, the evidence
presented relies purely on graphical representations and lacks a formal statisti-
cal test for significance which in turn leaves the final result open for debate. It
should be noted that the KSPA test is an extension of the KS statistic for compar-
ing between the predictive accuracy of two data sets. At present the KS statistic
is used for the purposes of distinguishing between the distributions of data and
this research presents an additional use of this statistic which is supported by
both simulation studies and applications to empirical data.

The beauty of the proposed KSPA test is that it not only enables distinguish-

ing between the distribution of forecasts from two models, but also enables
to determine whether the model with the lowest error also reports the lowest
stochastic error in comparison to the alternate model. Moreover, this test is not
affected by the potential autocorrelation that may be present in forecast errors
which is yet another advantage. The ability of exploiting the KSPA test for de-
termining the model with the lowest stochastic error stems from the work of
literature on stochastic dominance and as such deserves to be noted. Whilst the
consideration of stochastic dominance in forecasting literature is novel, as noted
in Horváth et al. (2006) stochastic dominance is widely used in econometric and
actuarial literature and is therefore a well established and recognized concept.
The use of KS tests for first and second order stochastic dominance dates back
to the work of McFadden (1989) where the author considered KS tests with in-
dependent samples with equal number of observations. Moreover, as the KS test
compares each point of the c.d.f. (Barrett and Donald, 2003; McFadden, 1989) it
has the potential of being a consistent test which considers all of the restrictions
imposed by stochastic dominance (Barrett and Donald, 2003).

The nature of the proposed KSPA test is such that it evaluates the differences
in the distribution of forecasting errors as opposed to relying on the mean dif-
ference in errors as is done in the DM approach. This in itself enables the KSPA
test to benefit from several advantages. Firstly, relying on the distribution of
errors enables the KS test to have more power than the DM test. This is because
the KSPA test essentially considers an infinite number of moments whilst the
DM test only tests the first moment which is popularly referred to as the mean.
Secondly, the presence of outliers can severely impact the DM test as the mean
is highly sensitive to outliers in data whereas the cumulative distribution func-
tion for errors are less affected. Thirdly, a test statistic which is concentrated
around a mean fails to account for the variation around the data. For example,
it is possible to have two populations with identical means and yet these two

populations would not really be identical if the variation around the mean is not the same. By considering the distribution of the data as is done via the proposed KSPA test, we are able to study and obtain a richer understanding of the underlying characteristics which in turn enables a more efficient and accurate decision.

The remainder of this chapter is organized as follows. The section which follows presents the theoretical foundation underlying the proposed statistical test for comparing between forecasting models. Section 6.3 is dedicated to the results from the simulation study which compares the size and power properties of both the KSPA and modified DM tests for different sample sizes and forecasting horizons. Section 6.4 presents empirical evidence from applications to real data where the performance of the KSPA test is compared alongside the modified DM test, and conclusions relating to this chapter are drawn in Section 6.5.

3.2   Theoretical Foundation

This section is dedicated to briefly introducing the theory underlying the Kolmogorov-Smirnov test which is followed by the introduction of the hypothesis for the two-sided and one-sided KS tests which are of interest to this research. Thereafter, the KSPA test is presented for distinguishing between the distribution of forecasts errors and identifying the model with the lower stochastic error. The first part of the KSPA test, which is the two-sample two-sided KSPA test, aims at identifying a statistically significant difference between the distribution of two forecast errors (and thereby comparing the predictive accuracy of forecasts). The second part, which is the two-sample one-sided KSPA test aims at ascertaining whether the forecast with the lowest error according to some loss function also has a stochastically smaller error in comparison to the competing forecast

(and thereby enables the comparison of the predictive accuracy of forecasts).

3.2.1   The Kolmogorov-Smirnov (KS) Test

The cumulative distribution function (c.d.f.) is an integral component of the KS test. As such, let us begin by defining the c.d.f., $F(x)$ for a random variable X. The c.d.f of X is denoted as:

$$F(x) = P(X \leq x), \tag{3.1}$$

where x includes a set of possible values for the random variable X. In brief, the c.d.f. shows the probability of X taking on a value less than or equal to x. The next step is to obtain the empirical c.d.f. This is because the one sample KS test (which is introduced below) aims at comparing the theoretical c.d.f. with an empirical c.d.f., whereby the latter is an approximation for the former. The empirical c.d.f. can be defined as:

$$F_n(x) = P_n(X \leq x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x), \tag{3.2}$$

where n is the number of observations, and I is an indicator function such that I equals 1 if $X_i \leq x$ and 0 otherwise. According to DeGroot and Schervish (2012), as implied by the law of large numbers, for any fixed point $x \; \varepsilon \; \mathbb{R}$, the proportion of the sample contained in the set $(-\infty, x]$ approximates the probability of this set as:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x) \Rightarrow \mathbb{E}I(X \leq x) = F(x), \tag{3.3}$$

where $\mathbb{E}$ represents the expectation.

Then, the one sample Kolmogorov–Smirnov statistic for any given $F(x)$ can

be calculated as

$$D_n = \max_x |F_n(x) - F(x)|, \tag{3.4}$$

where max denotes the maximum of the set of distances. Note that the one sample KS test in Equation (6.4) compares the empirical c.d.f. with a theoretical c.d.f. However, presented next is the two sample KS test statistic which is of direct relevance to the proposed KSPA test. In contrast to the one sample KS test, the two sample KS test compares the empirical c.d.f.'s of two random variables in order to find out whether both random variables share an identical distribution, or whether they come from different distributions. Assuming two random variables X and Y, the two sample KS test statistic will be

$$D_{n_1,n_2} = \max_x |F_{X,n_1}(x) - F_{Y,n_2}(x)|. \tag{3.5}$$

Next, we introduce the hypothesis which are relevant for the proposed KSPA test. Let us begin by presenting the hypothesis for the two-sided KS test. Let X and Y be two random variables with c.d.f.'s F_X and F_Y, respectively. Then, a two sample, two-sided KS test will test the hypothesis that both c.d.f.'s have an identical distribution, and the resulting null and alternate hypothesis can be expressed as:

$$H_0 : F_X(z) \equiv F_Y(z) \qquad \forall z \in \mathbb{Z}, H_1 : F_X(z) \neq F_Y(z), \text{ for some } z \in \mathbb{Z}. \tag{3.6}$$

In simple terms, the null hypothesis in Equation (5.6) states that both X and Y share an identical distribution whilst the alternate hypothesis states that X and Y do not share the same distribution.

Finally, the hypothesis for the two sample one-sided KS test which is also

known as the one-sided test of stochastic dominance is presented as in McFadden (1989):

$$H_0 : F_X(z) \leq F_Y(z) \qquad \forall z \in \mathbb{Z}, H_1 : F_X(z) > F_Y(z), \text{ for some } z \in \mathbb{Z}. \quad (3.7)$$

The important point to note here is that the alternate hypothesis in Equation (6.9) states that the c.d.f. of X lies above and to the left of the c.d.f. of Y, which in turn means that X has a lower stochastic error than Y. Note that in our case we consider X and Y in absolute or squared terms for example.

As with all tests, the decision making process requires the calculation of the probability value. For the KS test, there are various formulas for calculating the p-value, each with its own advantages and limitations. See for example, Birnbaum and Tingey (1951); Marsaglia et al. (2003) and Simard and L'Ecuyer (2011). The KSPA test relies on the formulae used in Simard and L'Ecuyer (2011) to calculate the p-values for both two-sided and one-sided KS tests. Introduced below are the two-sided and one-sided KSPA tests which are based on the foundations of the KS test which has been concisely explained above.

3.2.2 Testing for Statistically Significant Differences between the Distribution of Two Sets of Forecast Errors

The aim here is to exploit the two sample two-sided KS test (which is referred to as the two-sided KSPA test hereafter) to ascertain the existence of a statistically significant difference between the distributions of two forecast errors. Let us begin by defining forecast errors. Suppose we have a real valued, non zero time series $Y_N = (y_1, \ldots, y_t, \ldots y_N)$ of sufficient length N. Y_N is divided into two parts, i.e., training set and test set such that $Y_1 = (y_1, \ldots, y_t)$ represents the training set

and $Y_2 = (y_{t+1}, \ldots, y_N)$ represents the test set. The observations in Y_1 are used to model the data whilst the observations in Y_2 are set aside for evaluating the forecasting accuracy of each model. Assume we use two forecasting techniques known as m_1 and m_2. A loss function $\mathscr{L}$ can be used to assess and compare between the out-of-sample forecast errors. Whilst there are varied options for $\mathscr{L}$, here we define $\mathscr{L}$ as:

$$\mathscr{L}(y_{i+h} - \hat{y}_{i+h}), \qquad (i = t, \ldots, N - h) \tag{3.8}$$

where $h \geq 1$ denotes the forecasting horizon, and $\hat{y}_{i+h}$ denotes the h-step ahead forecast of Y_i. If the forecast error is denoted by ε, then we have the expression

$$\varepsilon_{i+h} = y_{i+h} - \hat{y}_{i+h}. \tag{3.9}$$

In this case the forecast errors for Y_2, obtained using models m_1 and m_2 can be denoted by

$$\varepsilon_{i+h}^{m_1} = y_{i+h} - \hat{y}_{i+h}^{m_1}, \ \ \varepsilon_{i+h}^{m_2} = y_{i+h} - \hat{y}_{i+h}^{m_2}, \tag{3.10}$$

where $\varepsilon_{i+h}^{m_1}$ is the h-step ahead forecast errors generated from model m_1 and $\varepsilon_{i+h}^{m_2}$ is the h-step ahead forecast errors generated from model m_2. The most common loss functions consider errors in the form of absolute values or squared values (see for example, the MAPE and RMSE). As such, we can use either the absolute value of errors or squared errors when calculating the KSPA test depending on the loss function in use. Then, the absolute values and squared values of forecast errors can be calculated as

$$\varepsilon_{i+h}^{m_1} = |y_{i+h} - \hat{y}_{i+h}^{m_1}|, \ \ \varepsilon_{i+h}^{m_2} = |y_{i+h} - \hat{y}_{i+h}^{m_2}|. \tag{3.11}$$

$$\varepsilon_{i+h}^{m_1} = (y_{i+h} - \hat{y}_{i+h}^{m_1})^2, \quad \varepsilon_{i+h}^{m_2} = (y_{i+h} - \hat{y}_{i+h}^{m_2})^2. \tag{3.12}$$

The forecast errors in (3.11) or (3.12) are inputs into the KSPA test for determining the existence of a statistically significant difference in the distribution of forecasts from models m_1 and m_2. As the requirement is to test the distribution between two samples of forecast errors, the two sample two-sided KSPA test statistic can be calculated as:

$$D_{i,i+h} = \max_x |F_{\varepsilon_{i+h}^{m_1}}(x) - F_{\varepsilon_{i+h}^{m_2}}(x)|, \tag{3.13}$$

where $F_{\varepsilon_{i+h}^{m_1}}(x)$ and $F_{\varepsilon_{i+h}^{m_2}}(x)$ denote the empirical c.d.f.'s for the forecast errors from two different models.

Accordingly, in terms of forecast errors, the two-sided KSPA test hypothesis can be approximately represented as follows; where $\varepsilon_{i+h}^{m_1}$ and $\varepsilon_{i+h}^{m_2}$ are the absolute or squared forecast errors from two forecasting models m_1 and m_2 with unknown continuous empirical c.d.f's, the two-sided KSPA test will test the hypothesis:

$$H_0 : F_{\varepsilon_{i+h}^{m_1}}(z) \equiv F_{\varepsilon_{i+h}^{m_2}}(z), H_1 : F_{\varepsilon_{i+h}^{m_1}}(z) \neq F_{\varepsilon_{i+h}^{m_2}}(z). \tag{3.14}$$

Then, if the observed significance value of the two-sample two-sided KSPA test statistic $D_{i,i+h}$ is less than α (which is usually considered at the 1%, 5% or 10% level), we reject the null hypothesis and accept the alternate which is that the forecast errors $\varepsilon_{i+h}^{m_1}$ and $\varepsilon_{i+h}^{m_2}$ do not share the same distribution. In such circumstances we are able to conclude with 1-α confidence that there exists a statistically significant difference between the distribution of forecasts provided by models m_1 and m_2, and thereby conclude the existence of a statistically significant difference between the two forecasts based on the two-sided KSPA test.

3.2.3 Testing for the Lower Stochastic Error

The aim of the two sample one-sided KS test (referred to as the one-sided KSPA test hereafter) is to identify whether the model which reports the lowest error based on some loss function also reports a stochastically smaller error in comparison to the alternate model. The usefulness of the one-sided KSPA test in distinguishing between the predictive accuracy of forecasts is most apparent in circumstances where forecasts from two models may share an identical distribution with some degree of error (as otherwise this would mean the two forecasts are exactly the same), such that one model will clearly report a comparatively lower forecast error based on some loss function. In such instances, the two-sided KSPA test would fail to identify a statistically significant difference between the two forecasts, but the one-sided KSPA test has the ability of testing the out-of-sample forecasts further in order to identify whether the model with the lower error also reports a stochastically smaller error, and thereby test for the existence of a statistically significant difference between two forecasts.

In terms of forecast errors, the two-sample, one-sided KSPA test hypothesis can be approximately represented as follows. Once again, where $\varepsilon_{i+h}^{m_1}$ and $\varepsilon_{i+h}^{m_2}$ are the absolute or squared forecast errors from two forecasting models m_1 and m_2 with unknown continuous empirical c.d.f.'s, the two sample one-sided KSPA test will test the hypothesis:

$$H_0 : F_{\varepsilon_{i+h}^{m_1}}(z) \leq F_{\varepsilon_{i+h}^{m_2}}(z), H_1 : F_{\varepsilon_{i+h}^{m_1}}(z) > F_{\varepsilon_{i+h}^{m_2}}(z). \tag{3.15}$$

The acceptance of the alternate hypothesis in this case translates to the c.d.f. of forecast errors from model m_1 lying towards the left and above the c.d.f. of forecast errors from model m_2. More specifically the acceptance of the alternate hypothesis confirms that model m_1 reports a lower stochastic error than model m_2. Recall the relationship identified in Hassani et al. (2009) that if the c.d.f.

for absolute value of forecast errors from one model lies above and hence to the left of that for the other model, the model lying above had a lower stochastic error than the other model. The one-sided KSPA test evaluates this notion and provides a statistically valid foundation which was previously lacking.

3.3 Simulation Results

3.3.1 Size of the Test

The first part of the simulation study focuses mainly on the size properties of the proposed KSPA test. The actual size of nominal 10% level tests are estimated against a two-sided alternative as in Diebold and Mariano (1995) and Harvey et al. (1997), and the simulation study itself follows the exact process as in Harvey et al. (1997). This research has considered and reported the results from errors generated via various noise distributions, and as an example explained below is one of the processes, i.e. the process involved in the Gaussian white noise simulation. Independent standard normal white noise error series were simulated (e_{1t}, e_{2t}), $t = 1, 2, \ldots, n$, for various sample sizes n. Forecasts which cover both short and long run horizons, more specifically up to $h = 10$ were evaluated. As in Harvey et al. (1997), the information related to the simulated white noise error series were incorporated in the test statistics only in the case of $h = 1$. In order to enable comparison with the results in Harvey et al. (1997), the squared errors were considered, i.e. e_{1t}^2 and e_{2t}^2 over the entire simulation study. All simulation results reported are based on 10,000 replications and were programmed in R.

The results for the size properties are reported in Table 3.1. Firstly, as noted in Harvey et al. (1997) the modified DM test remains somewhat oversized as visible in the results shown in Table 3.1. Yet, the authors concluded this was

acceptable as the modified DM test results showed a major improvement over the previous version. Based on the results, we see that for the Guassian white noise errors which are directly comparable with the modified DM test results in Harvey et al. (1997), the proposed KSPA test is correctly sized across all sample sizes, both large and small, and across all horizons of up to 10 steps-ahead.

Unlike the results reported in Harvey et al. (1997), also shown here are the outcomes from the simulation study which considered Uniform distribution, Cauchy distribution, and heavy tailed errors. The Cauchy white noise distributions are likely in time series which are affected by catastrophic events. The heavy tailed distribution is a Student's t distribution with six degrees of freedom as considered and explained in Harvey et al. (1997). The findings from the simulation study indicates the superiority of the proposed KSPA test over the modified DM test in terms of being correctly sized across all sample sizes and all horizons even when faced with varying noise distributions.

It is noteworthy that the results reported in Table 3.1 represents a subset of results obtained from an extensive simulation study. Following the simulation study in Harvey et al. (1997), also considered are (1) contemporaneously correlated forecast errors with contemporaneous correlations of 0.5 and 0.9, and (2) autocorrelated forecast errors. The results were similar to what is reported for the other distributions in Table 3.1 as it continued to illustrate that the KSPA test is indeed correctly sized across all sample sizes and forecasting horizons. As such in order to save space these results are not reported here.

In summary, it is evident that in comparison to the modified DM test, the KSPA test shows major improvements not only across different forecasting horizons, but more importantly over small sample sizes. As noted in Harvey et al. (1997) it is the performance over small sample sizes that is of utmost importance to practitioners as in reality very large number of forecasts are not often available for comparison purposes, and the proposed KSPA test has proven to

be comparatively more accurate in this case with far better results.

3.3.2   Power of the Test

Next, the power of the modified DM test and the proposed KSPA test are considered. For this purpose, forecast errors were generated from different combinations of distributions which will certainly result in significantly different forecast errors, so that one can obtain an accurate evaluation of the power of the tests in relation to small and large sample sizes. The details of the combinations evaluated are explained in the footnotes of Table 3.2 which also reports the results. The power of the two tests were evaluated over the one-step horizon because power calculations are only valid if a given test is correctly sized, and the modified DM test suffers from problems of being oversized especially beyond $h = 1$ Harvey et al. (1997).

Once again, reported here is a subset of all results as the general conclusion remains similar. Firstly, it is evident that the KSPA test is more powerful than the modified DM test for both small and large sample sizes. Secondly, the KSPA test converging towards a power of 100% faster than the modified DM test in all cases

whereas on most instances the results (including those not reported here) showed that the modified DM test fails to converge to 100% over these sample sizes. The only exception being in the case of autocorrelated errors as in Case 3, skewed errors as in Case 4 or where forecast errors generated from a MA(1) process was compared against those generated from an AR(1) process.

In summary, the simulation study has shown that the proposed KSPA test is correctly sized across all sample sizes and forecasting horizons, and that it is more powerful than the DM test, and thereby proving its practicality and suitability as a complement statistical test for distinguishing between the predictive

Table 3.1 Percentage of rejections of the true null hypothesis of equal prediction mean squared errors for the Diebold-Mariano test and equal distribution of squared prediction errors for the KSPA test at nominal 10% level.

h	Error Distribution	Test	n=8	n=16	n=32	n=64	n=128	n=256	n=512
1	Gaussian	DM	8.4	9.6	9.7	10.1	9.9	10.4	10.6
	Gaussian	KSPA	8.6	9.4	8.9	9.6	8.4	9.4	8.6
	Uniform	KSPA	9.1	8.9	8.6	9.4	8.9	8.9	8.5
	Cauchy	KSPA	9.0	9.1	8.4	9.2	8.5	8.9	8.6
	Student's t	KSPA	8.5	9.4	9.3	9.5	9.0	8.7	8.6
2	Gaussian	DM	16.4	14.2	12.2	11.2	10.8	10.5	10.3
	Gaussian	KSPA	9.0	9.5	8.5	9.2	8.6	9.1	8.4
	Uniform	KSPA	9.1	9.4	8.9	9.8	8.8	9.2	8.8
	Cauchy	KSPA	9.3	9.5	9.0	9.3	8.8	9.4	9.0
	Student's t	KSPA	8.7	9.3	9.1	9.1	8.4	9.7	8.9
3	Gaussian	DM	18.1	18.5	14.3	12.2	10.7	10.8	10.9
	Gaussian	KSPA	8.6	9.6	8.7	9.2	8.7	9.1	9.1
	Uniform	KSPA	8.7	9.8	9.0	9.2	8.6	9.4	8.7
	Cauchy	KSPA	8.4	9.4	9.3	9.7	8.7	9.5	8.7
	Student's t	KSPA	8.2	9.7	8.8	9.5	8.9	9.1	8.6
4	Gaussian	DM	16.3	19.8	16.1	13.4	11.5	10.9	11.0
	Gaussian	KSPA	8.5	9.4	8.3	8.9	8.6	9.2	9.0
	Uniform	KSPA	8.7	9.6	8.6	9.2	9.4	9.6	9.1
	Cauchy	KSPA	8.4	9.4	9.0	9.4	9.6	9.7	8.7
	Student's t	KSPA	8.7	9.1	8.8	9.9	8.7	9.7	8.8
5	Gaussian	DM	12.9	19.9	17.8	14.9	12.2	11.1	11.0
	Gaussian	KSPA	8.4	9.4	8.9	9.4	8.3	9.7	8.3
	Uniform	KSPA	8.2	9.2	8.7	9.1	8.4	9.3	8.9
	Cauchy	KSPA	8.8	9.6	8.5	9.5	9.0	8.8	8.9
	Student's t	KSPA	8.4	9.3	9.1	9.9	9.1	9.6	8.6
6	Gaussian	DM	10.6	19.8	18.8	16.0	12.9	11.4	11.2
	Gaussian	KSPA	8.6	9.5	8.9	9.5	8.6	9.1	9.0
	Uniform	KSPA	8.7	9.4	8.8	9.1	8.4	9.2	8.3
	Cauchy	KSPA	8.9	9.8	9.1	9.9	8.5	9.2	8.6
	Student's t	KSPA	8.7	9.3	8.8	9.4	9.0	9.8	9.1
7	Gaussian	DM	9.9	18.2	19.5	16.8	13.6	11.6	11.4
	Gaussian	KSPA	8.6	9.5	9.3	8.9	8.8	9.3	9.0
	Uniform	KSPA	8.4	9.0	8.7	9.9	9.0	9.1	8.7
	Cauchy	KSPA	8.5	9.2	8.7	9.1	9.0	9.4	8.9
	Student's t	KSPA	8.8	9.1	9.0	9.0	8.6	8.8	9.2
8	Gaussian	DM	-	17.4	20.2	18.0	13.8	11.9	11.4
	Gaussian	KSPA	-	9.3	8.6	9.1	8.5	9.5	8.7
	Uniform	KSPA	-	9.5	8.7	9.8	9.0	9.7	8.7
	Cauchy	KSPA	-	9.5	8.3	9.2	8.8	8.9	8.9
	Student's t	KSPA	-	9.7	8.3	9.6	8.6	9.1	9.1
9	Gaussian	DM	-	15.1	20.2	19.0	14.7	12.4	11.6
	Gaussian	KSPA	-	9.5	8.6	9.2	8.5	9.4	8.8
	Uniform	KSPA	-	9.4	9.0	9.7	8.0	9.5	8.9
	Cauchy	KSPA	-	9.8	8.6	8.9	8.6	9.4	8.8
	Student's t	KSPA	-	9.1	8.6	9.2	8.9	9.6	9.0
10	Gaussian	DM	-	14.0	20.2	19.1	15.1	12.6	11.8
	Gaussian	KSPA	-	9.2	8.9	9.3	8.7	9.7	9.0
	Uniform	KSPA	-	9.2	8.7	9.8	8.7	9.1	9.4
	Cauchy	KSPA	-	9.2	8.8	9.7	9.1	9.5	9.3
	Student's t	KSPA	-	9.3	8.8	9.0	8.7	9.1	8.6

Note: The DM test results relate to modified DM test and were extracted from Table 1 in Harvey et al. (1997).

accuracy of forecasts.

Table 3.2 Percentage of rejections of the false null hypothesis of equal one-step prediction mean squared errors for the Diebold-Mariano test and equal one-step distribution of squared prediction errors for the KSPA test at nominal 10% level.

Combinations	Test	n=8	n=16	n=32	n=64	n=128	n=256	n=512
Case 1	DM	7.3	17.5	31.9	37.3	39.3	40.3	40.9
	KSPA	19.6	35.8	61.0	91.7	99.9	100.0	100.0
Case 2	DM	5.2	13.4	26.5	35.4	39.5	41.0	40.8
	KSPA	15.9	25.8	42.0	75.3	97.6	100.0	100.0
Case 3	DM	59.3	96.0	99.7	100.0	100.0	100.0	100.0
	KSPA	65.1	92.0	100.0	100.0	100.0	100.0	100.0
Case 4	DM	91.6	99.7	100.0	100.0	100.0	100.0	100.0
	KSPA	97.3	100.0	100.0	100.0	100.0	100.0	100.0

Note: Case 1: Compares errors from a Cauchy distribution with mean 0 and standard deviation 1 against errors from a $N(0,1)$ distribution. Case 2: Compares errors from a Student's t distribution with 6 d.f. against errors from a Cauchy distribution. Case 3: Compares errors from $N(0,1)$ against autocorrelated errors. Case 4: Compares errors from a skewed χ^2 distribution with 3 d.f. against errors from a χ^2 distribution with 10 d.f.

3.4 Empirical Evidence

Following the simulation study which illustrated the superiority of the proposed KSPA test in terms of being correctly sized and more powerful than the modified DM test, discussed in this section is the use of the KSPA test for several real world applications. Note that all applications here use the RMSE as the loss function, and therefore the KSPA test like the DM test relies on squared errors in all instances. These real world applications have been carefully selected to illustrate that: (i) The KSPA test can accurately perform the same task as the modified DM test in practice when faced with real data. (ii) Both two-sided and one-sided KSPA tests can be of benefit in practice. (iii) The KSPA test is applicable where the modified DM test cannot be applied. (iv) The KSPA test can handle both small and large sample sizes. (v) The KSPA test is suitable across different forecasting horizons. (vi) The KSPA test is not affected by the generation of forecast errors from either parametric or non-parametric models.

3.4.1 Scenario 1: Tourism Series

The rationale for this application is to mainly show that the KSPA test can perform the same task as the modified DM test in practice in addition to showing its performance when faced with a relatively large number of out-of-sample forecast errors. We consider testing forecasts from two models, i.e. Singular Spectrum Analysis (SSA) which is non-parametric and ARIMA (parametric) for a statistically significant difference in terms of providing $h = 1$ step ahead forecasts for total U.S. tourist arrivals[2]. This monthly data set is used in Chapter 3 where more information pertaining to the data is available, and the related forecasts of 69 observations were extracted from that application. Figure 3.1 shows the out-of-sample forecasts, distribution of errors and the empirical c.d.f. for U.S. tourist arrivals obtained via SSA and ARIMA models. Based on the forecasts figure alone one is not able to determine whether there exists a statistically significant difference between the forecasts from SSA and ARIMA. As such, we then look to the distribution of the squared forecast errors from ARIMA and SSA which can be seen in Figure 3.1 (middle). However, without a formal statistical test it is not possible to determine whether there exists a statistically significant difference between the distribution of these errors. Thirdly, we look at the empirical c.d.f.'s shown in Figure 3.1 (right) to identify if one model does indeed provide a lower stochastic error than the other model as suggested in Hassani et al. (2009). In this case it is clear that based on the empirical c.d.f., it appears that the out-of-sample forecasts from SSA provide a lower stochastic error than the out-of-sample forecasts from ARIMA. However, as mentioned in the introduction this conclusion is open to debate as it lacks a mandatory statistical test.

When we calculate the RRMSE statistic, it shows that the forecasts obtained

---

[2]Data source: http://travel.trade.gov/research/monthly/arrivals/

from the SSA model are 60% better than the forecasts obtained via the ARIMA model. Accordingly one would expect a statistically significant difference between the forecasts of SSA and ARIMA. Both the modified DM and KSPA tests are applied. The results are reported in Table 3.3. In this case, the modified DM test correctly identifies that there exists a statistically significant difference between the forecasts from SSA and ARIMA. In terms of the newly proposed KSPA test, firstly the two-sided KSPA test confirms that there is indeed a statistically significant difference between the distribution of forecast errors from SSA and ARIMA for U.S. tourist arrivals at $h = 1$ step ahead, and thereby confirms the existence of a statistically significant difference between the two forecasts. Next, the one-sided KSPA test is applied to find out whether SSA forecasts (which has the lower RMSE) reports a lower stochastic error than ARIMA forecasts. The one-sided KSPA test confirms that SSA does in fact provide forecasts which report a lower stochastic error than the ARIMA model as suggested by the empirical c.d.f.'s in Figure 3.1 (right), and provides supplementary evidence to the conclusion from the two-sided KSPA test for the existence of a statistically significant difference between the two forecasts. The results from the modified DM test and KSPA tests are significant at a 95% confidence level.

Table 3.3 Evaluating $h = 1$ step ahead forecasts for U.S. tourist arrivals.

Test	Two-sided (p-value)	One-sided (p-value)
Modified DM	<0.01*	N/A
KSPA	<0.01*	<0.01*

Note:* indicates results are statistically significant based on a p-value of 0.05. N/A refers to not applicable as a directly comparable alternative form of the DM test was not available in the code used.

3.4.2 Scenario 2: Accidental Deaths Series

The main reason to present this next application is to show how the KSPA test can overcome a limitation of the modified DM test. The well known U.S. death

series includes monthly data with 78 observations recorded between January 1973-June 1979 and has been used widely in previous time series analysis and forecasting applications (see for example, (Brockwell and Davis, 2002; Hassani, 2007; Hassani et al., 2014)). This application follows a similar forecasting approach to that reported in Hassani (2007). The application looks at forecasting the last 12 points of the death series such that the first forecast point represents the horizon of $h = 1$, the second forecast point represents $h = 2$ and so on, up until the final forecast point which represents the $h = 12$ steps ahead forecast. As explained in Section 6.1, the modified DM test cannot be used in such scenarios where the out-of-sample forecast errors relate to various horizons within a single forecasting exercise as the sum of the covariance will equate to zero. However, it is possible to rely on the original DM test (i.e. without considering the covariance effect) in such scenarios, but it is not advisable owing to the many limitations of the original DM test as identified in Harvey et al. (1997). The forecasts are obtained via the parametric ARIMA model and a non-parametric Neural Networks (NN) model, and the ARIMA forecasts report a lower RMSE.

Figure 3.2 shows the out-of-sample forecasts, distribution of errors and the empirical c.d.f. for the U.S. death series obtained via ARIMA and NN models. In this case based on the empirical c.d.f. we are able to state that the ARIMA forecasts report a stochastically smaller error than the NN forecasts. The two-sided KSPA test can be used to test for statistically significant differences between the two forecasts whilst the one-sided KSPA test can be exploited to provide statistical evidence for the claim based on Hassani et al. (2009). The resulting output from the KSPA and original DM tests are reported in Table 3.4. Initially, the two-sided KSPA test confirms that there is indeed a statistically significant difference between the distribution of forecast errors from ARIMA and NN at a 95% confidence level. Secondly, the one-sided KSPA test confirms that ARIMA does in fact provide forecasts which report a lower stochastic error

than forecasts from the NN model as suggested by the empirical c.d.f.'s in Figure 3.2 (right). Note that whilst the original DM test too proves the existence of a statistically significant difference between the two forecasts, the two-sided KSPA test reports a lower p-value than the original DM test.

Table 3.4 Evaluating $h = 1, \ldots, h = 12$ steps ahead forecasts for the U.S. death series.

Test	Two-sided (p-value)	Greater (p-value)
DM	0.04*	N/A
Modified DM	N/A	N/A
KSPA	0.03*	0.02*

Note:* indicates results are statistically significant based on a p-value of 0.05. N/A refers to not applicable as a directly comparable alternative form of the DM test was not available in the code used.

3.4.3 Scenario 3: Trade Series

Finally, the purpose of this empirical example is to show the superiority of the KSPA test over the modified DM test, and also to show how the one-sided KSPA test is useful when the two-sided KSPA test fails at showing a statistically significant difference between two forecasts. In this application we consider forecasts for monthly U.S. imports[3] between March 2011-December 2011 (10 observations) at $h = 3$ steps ahead using ETS and SSA which are both non-parametric techniques. This data set was recently used in Silva and Hassani (2015) and the forecasts considered here are those generated in that study. This is another example of a scenario with a small sample size i.e. $n = 10$. Figure 3.3 shows the out-of-sample forecasts, distribution of errors and the empirical c.d.f. of errors obtained via ETS and SSA. Here, unlike on previous occasions, based on this forecast figure alone one is able to see that there exists a significant difference between the forecasts from both models. However, it cannot be verified in the absence of statistical evidence. The distribution of out-of-sample fore-

---

[3]Data source: http://www.bea.gov/international/index.htm.

cast errors are shown in Figure 3.3 (middle) and the resulting empirical c.d.f.
for the squared forecast errors are also presented (right). Based on the empir-
ical c.d.f. in Figure 3.3 we can see that except for three points, at every other
observation, the forecasts from SSA appear to report a smaller stochastic error
than the forecasts from the ETS model (according to the inference in Hassani
et al. (2009)). Once again, relying solely on this empirical c.d.f. in Figure 3.3
(right) will only result in conclusions which are debatable. The RRMSE crite-
rion shows that forecasts from the SSA model are 54% better than those from
the ETS model. The expectation would be that such a significant gain reported
through the RRMSE will appear as statistically significant. In order to confirm
the expectations we apply both modified DM and KSPA tests.

The results from the two tests are reported in Table 3.5. Based on the modi-
fied DM test we are inclined to conclude that there exists no statistically signifi-
cant difference between the forecasts obtained via ETS and SSA. The modified
DM test statistic reports a p-value which exceeds 10% thus leading to the ac-
ceptance of the null hypothesis. Likewise, the two-sided KSPA test suggests
that forecast errors obtained via ETS and SSA share an identical distribution.
Accordingly, there is no sufficient evidence based on the modified DM test and
the two-sided KSPA test for a statistically significant difference between the
two forecasts. However, this is where the one-sided KSPA test becomes ex-
ceedingly useful. We know based on the RRMSE criterion that forecasts from
SSA report a lower RMSE than forecasts from ETS. As such, we can use the
one-sided KSPA test to find out whether the SSA forecast which reports a lower
error based on the RMSE loss function also reports the lowest stochastic error in
comparison to the ETS forecast. Accordingly there is sufficient evidence based
on the one-sided KSPA test at the 10% significance level to conclude that SSA
forecasts report a lower stochastic error than forecasts from ETS. Thereby, one
can confirm the existence of a statistically significant difference between the two

forecasts which is expected given that forecasts from SSA are 54% better than forecasts from ETS according to the RRMSE criterion.

The results from this case not only show the advantage of the one-sided KSPA test, but also proves that the graph of one c.d.f. need not lie 'strictly' above the graph of another c.d.f. as suggested in Hassani et al. (2009) in order for one model to report a stochastically smaller error than the alternate model. It is clear that if a larger proportion of the c.d.f. of errors from one model lies above the c.d.f. of errors from another model, then the KSPA test is able to accurately show that one model reports a lower stochastic error than the other, and thereby pick up a statistically significant difference between the forecasts from two models.

Table 3.5 Evaluating $h = 3$ step ahead forecasts for U.S. imports.

Test	Two-sided (p-value)	Greater (p-value)
Modified DM	0.30	N/A
KSPA	0.17	0.08*

Note:* indicates results are statistically significant based on a p-value of 0.10. N/A refers to not applicable as a directly comparable alternative form of the DM test was not available in the code used.

3.5   Discussion

Developing on the ideas presented in Hassani et al. (2009) and Hassani et al. (2013b) with respect to using an empirical c.d.f. for determining whether the forecast errors from one model are stochastically smaller than those obtained from a competing model, introduced in this chapter is a complement statistical test for distinguishing between the predictive accuracy of forecasts. The proposed non-parametric Kolmogorov-Smirnov Predictive Accuracy (KSPA) test serves two purposes via the two-sided KSPA test and the one-sided KSPA test. A simulation study is called upon to evaluate the efficiency and robustness of

the KSPA test which is followed by an application to real data. The need for the
KSPA test is further evidenced by limitations of the DM test in relation to issues
in sample size or inherent assumptions which have been left invalidated in the
face of recent findings.

Through the simulation study, the KSPA test is directly compared with the
widely accepted modified DM test. In order to enable a meaningful compari-
son, the same distributions as used in Harvey et al. (1997) for their simulation
study are considered here. The simulation results provide a clear indication that
the proposed KSPA test is more robust than the DM test especially when the
number of out-of-sample forecast errors available for comparison purposes are
considerably small.

Also considered are applications to real data which capture forecasts from
different cases in real world applications for validating the proposed KSPA test,
and compare the results against those obtained via the modified DM test. As
expected, it was observed that when the number of observations are small the
KSPA test is able to accurately identify a statistically significant difference be-
tween forecasts whilst the modified DM test fails. Furthermore, through another
scenario in real world applications it is shown that the KSPA test can be applied
in forecasting exercises where the modified DM test is not applicable. In addi-
tion, another scenario is used to show that the two variations of the KSPA test
can be extremely useful in practice.

Yet another advantage in the proposed KSPA test is that given its nature,
which is to compare the empirical c.d.f. of errors from two forecasting models,
one is able to compare both parametrically estimated model-based forecasts and
survey-based forecasts with no restrictions on whether these models are nested
or non-nested. This is because regardless of the model used, a forecast error
will always be calculated as the actual value minus the predicted value, and the
proposed KSPA test will compare the distribution of these errors to differentiate

between them. In addition, as the KSPA test is non-parametric it is not dependent on any assumptions relating to the properties of the underlying errors which is also advantageous in practice.

In conclusion, the KSPA test has shown promising results in comparison to the modified DM test and is presented as a viable alternative for comparing between the predictive accuracy of forecasts. The non-parametric nature of the test enables one to overcome issues with the assumptions underlying the DM test which have recently been proven void (see for example, (Hassani, 2010; Hassani et al., 2012a)). Additionally, this research provides statistical validity to the ideas presented in Hassani et al. (2009) and Hassani et al. (2013b) whilst showing the relevance and applicability of the KSPA test via simulations and applications to real data. Future research relating to this test continues to ascertain whether there is a possibility of extending the use of the KSPA test to enable comparisons between more than two forecasts as this would add more value to its practical use.
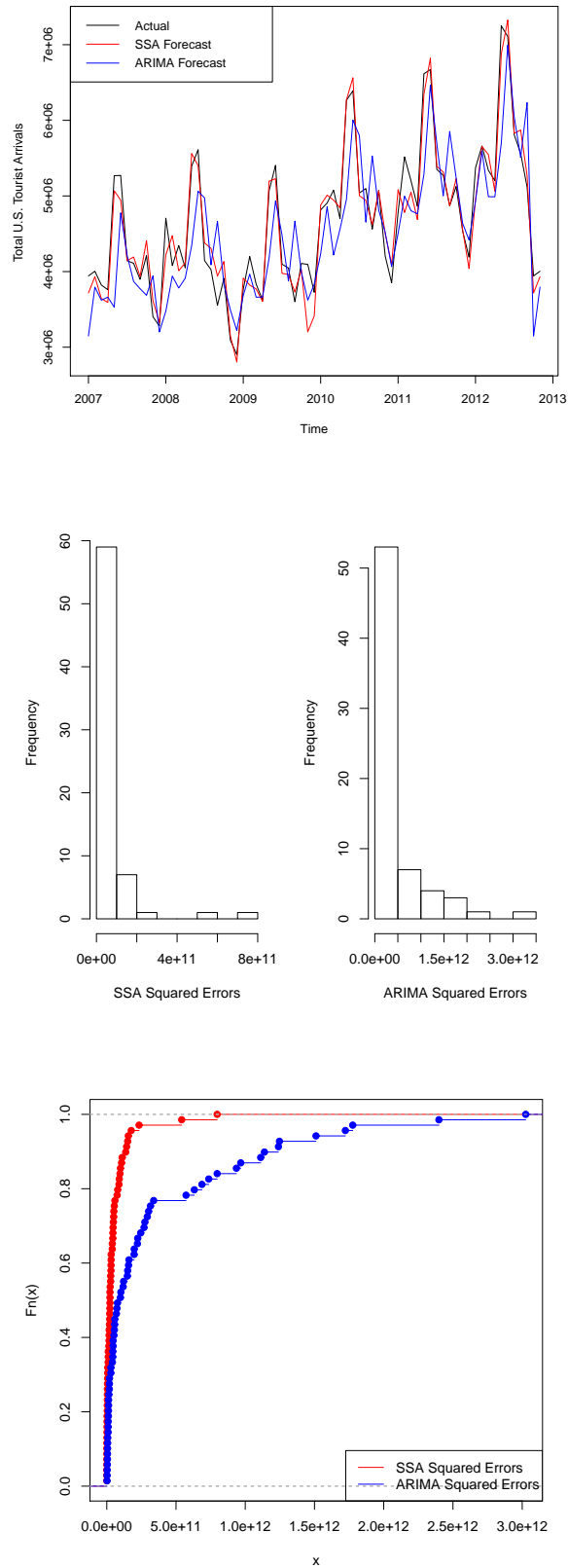
Fig. 3.1 U.S. Tourist arrivals forecast, distribution of errors and empirical c.d.f. of errors.
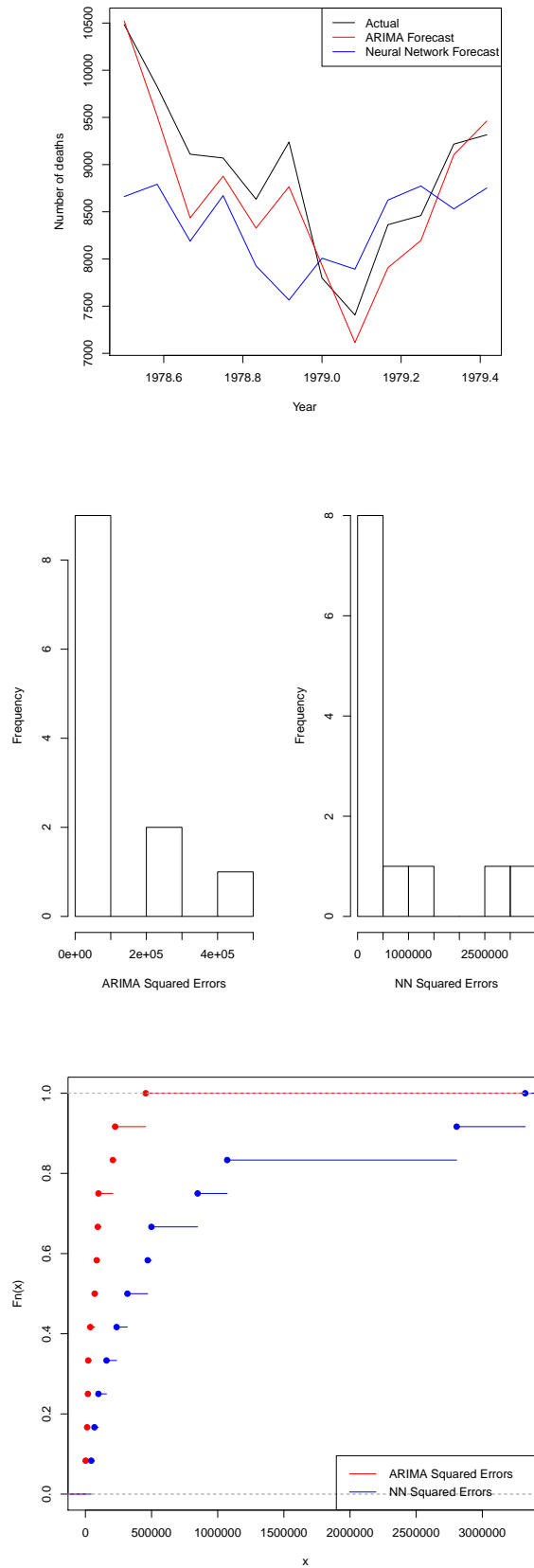
Fig. 3.2 U.S. death series forecast, distribution of errors and empirical c.d.f. of errors.
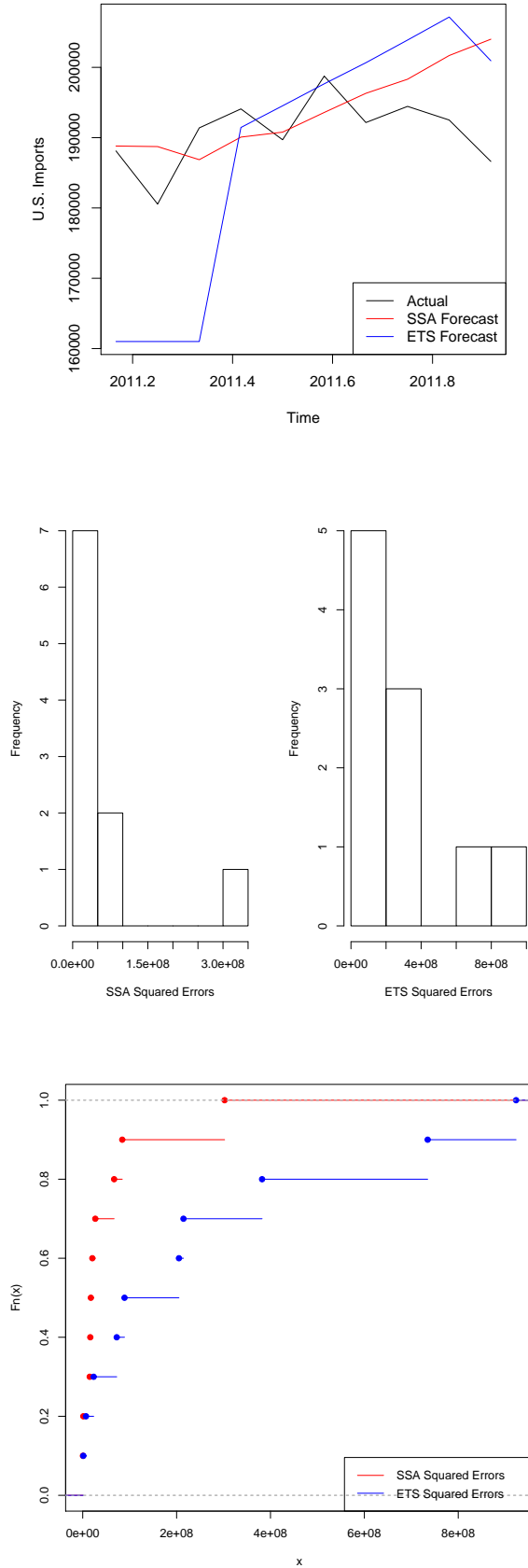
Fig. 3.3 U.S. imports forecast, distribution of errors and empirical c.d.f. of errors.

Chapter 4

Exploiting the Forecastability of Forecasts

This Chapter is aimed at the introduction of a new theoretical framework for exploiting the forecastability of forecasts which is also the second contribution of this thesis. The chapter begins with a concise introduction which discusses the need and significance of the proposed theory and is then followed by the introduction of the theory itself. The chapter also includes applications to real data.

4.1 Introduction

Forecasting continues to remain a top priority for planning and decision making in any given company, industry or economy. Whilst the ever increasing volatility and uncertainty in markets has further augmented the difficulty associated with obtaining accurate forecasts, the emergence of Big Data on the other hand has provided new insights and opportunities for improving and enhancing the accuracy of forecasts for any given variable. In the past, univariate forecasting (i.e. for example using historical monthly GDP forecasts for obtaining future

monthly GDP forecasts) has been the most popular norm with a wide range of applications. The results from such efforts have been productive, but the introduction and applications of multivariate forecasting approaches have provided far greater outcomes with increased accuracy levels.

Governments, practitioners, researchers and private organizations publish a variety of forecasts each year. Such forecasts are generally computed using multivariate models and are widely used in decision making processes given the considerably high level of anticipated forecast accuracy. The classical multivariate methods consider modelling multiple information pertaining to the same time period or with a time lag into the past. However, the focus of this research goes beyond the classical approaches and considers devising a novel theoretical framework for exploiting information pertaining to the future for further enhancing the accuracy of such predictions.

The aim of this chapter is to introduce a novel theoretical development which seeks to exploit the information contained in published forecasts (which represent data with a time lag into the future) for generating a new and improved (comparatively more accurate) forecast by taking advantage of the MSSA technique's capability at modelling time series with different series lengths. In brief, the proposed multivariate theoretical development seeks to exploit the forecastability of forecasts by considering not only official and professional forecasts, but also forecasts obtained via other time series models. As mentioned previously, the SSA technique has both univariate and multivariate forecasting capabilities along with two main forecasting options known as the Recurrent and Vector approach. The MSSA technique further divides into HMSSA and VMSSA.

Here in lies the beauty of MSSA in comparison to other multivariate forecasting methods. The HMSSA algorithm enables one to model and forecast time series with the same length whereas the VMSSA algorithm enables modelling and forecasting using time series with different lengths. This research takes ad-

vantage of this unique modelling capability of VMSSA and develops a theory for exploiting the forecastability of forecasts by modelling data with a time lag into the future. In brief, the main idea is to evaluate whether it is possible to exploit VMSSA by making use of historical data for a given variable in combination with either an official or professional forecast to improve upon the existing forecast's accuracy. The theory is evaluated with real data which considers not only official and professional forecasts, but also forecasts generated via other time series models. The main objective is to ascertain whether the new theoretical proposition enables to generate a forecast which can outperform the official forecast accuracy (or professional forecast or forecasts from another model as relevant). In addition, the SSA-R and SSA-V forecasts are also considered as benchmarks. Given the introductory nature of this theoretical concept the one of the most important points to initially evaluate is whether the proposed MSSA approach can successfully outperform the SSA benchmarks. This is because if it cannot do so, then there is no sufficient evidence for researching further into improving this theory further.

In practice it is possible that during the model training and testing procedure we would experience certain models which are capable of providing forecasts which outperforms forecasts from SSA. Likewise, official and professional forecasts are very likely to be extremely accurate given the wide ranging information that has been considered in arriving at the said predictions. Via the proposed theoretical development, this research attempts to exploit such superior forecasts from either other models, official or professional forecasts in order to improve the existing forecasts by modelling with VMSSA. This research also marks the first ever attempt at exploiting information contained within official or professional forecasts for generating a more accurate forecast. Two important points to note are that; firstly, the usual multivariate modelling problem involves using two different time series and extracting any useful information for improving

the accuracy of forecasts for both variables or one of the two variables. However, considered here is the same variable and a forecast for that same variable to generate a new set of forecasts which can provide better accuracy. Secondly, not all multivariate forecasting models can exploit this new idea as they cannot model when faced with different series lengths which is a major advantage and flexibility of the MSSA technique. Given the novel nature of this proposed approach it is important to note that there is no published academic literature which seeks to exploit the forecastability of forecasts by re-modelling a forecast which represents data with a time lag into the future in combination with historical data for developing a new and improved forecast. In particular it should be noted that this proposed methodology does not fall in line with literature on forecast combining which has been developed over the years.

4.2   Theoretical Development

Assume that we have a monthly time series $Y_N^{(1)}$ on length N, and further auxiliary information of a h-step ahead forecast for that series contained in Ω. Note that $Y_N^{(1)}$ and Ω are time series with different series lengths as shown below. The data in Ω can represent an official or professional forecast for $Y_N^{(1)}$, achievable using any method of forecasting. The hypothesis is that, provided the information contained in Ω is of some level of accuracy, then we can model this information alongside historical information for that same variable in a MSSA framework to develop an all new forecast for $Y_N^{(1)}$. The MSSA technique which can model time series with different lengths allows the exploitation of any auxiliary information contained within Ω and uses this in combination with the historical information found in $Y_N^{(1)}$, to produce a new forecast which can outperform the forecasts obtained by only $Y_N^{(1)}$ in terms of accuracy. For explanation purposes, let us assume $Y_N^{(1)}$ is the actual monthly inflation values and Ω is the h-step

ahead forecast for inflation such that:

$$Y_N^{(1)} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \omega_{N+1} \\ \omega_{N+2} \\ \vdots \\ \omega_{N+h} \end{pmatrix}, \tag{4.1}$$

A new time series can be constructed by incorporating the forecasted values with the actual values such that, $Y_{N+h}^{(2)} = (Y_N^{(1)}, \Omega)$.

$$Y_{N+h}^{(2)} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \\ \omega_{N+1} \\ \vdots \\ \omega_{N+h} \end{pmatrix}. \tag{4.2}$$

and the following corresponding trajectory matrix $\mathbf{X}^{(2)} = [X_1, ..., X_K, ..., X_{K+h}]$ can be computed, such that

$$\mathbf{X}^{(2)} = (x_{ij})_{i,j=1}^{L,K+h} = \begin{pmatrix} y_1 & y_2 & \cdots & y_K & y_{K+1} & \cdots & \omega_{K+h} \\ y_2 & y_3 & \cdots & y_{K+1} & y_{K+2} & \cdots & \omega_{K+h+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N & \omega_{N+1} & \cdots & \omega_{N+h} \end{pmatrix}. \tag{4.3}$$

Recall that the hypothesis states, if the information contained in Ω is accurate, then it is possible to exploit this information in a MSSA framework to obtain a new forecast that can outperform the accuracy of using only $Y_N^{(1)}$. Similar to the process in SSA, we can define the trajectory matrices $\mathbf{X}^{(i)}$ ($i = 1, 2$) of the one-dimensional time series $Y_{N_i}^{(i)}$ ($i = 1, 2$) with different series length. Thus, ap-

plying the above procedure to each series separately provides 2 different $L_i \times K_i$ trajectory matrices $\mathbf{X}^{(i)}$ $(i = 1, 2)$.

After embedding we organise a block Hankel matrix. According to Hassani and Mahmoudvand (2013) the MSSA approach has two main variations based on how one organizes the trajectory matrix $\mathbf{X}$. These are referred to as Horizontal MSSA (HMSSA) and Vertical MSSA (VMSSA). Here, we consider the MSSA approach in a vertical form, however there are some restrictions in selection the values of K and it is required to have $K_1 = K_2 = K$. Accordingly, the VMSSA approach enables us to have various window length L_i and different series length N_i, but as we mentioned above similar K_i for all series. The block Hankel trajectory matrix can then be defined as

$$\mathbf{X}_V = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}, \tag{4.4}$$

where, $\mathbf{X}_V$ indicates that the output of the embedding step is in a *vertical* form. Next, we obtain the SVD of $\mathbf{X}_V$. Denote $\lambda_{V_1}, \ldots, \lambda_{V_{L_{sum}}}$ as the eigenvalues of $\mathbf{X}_V \mathbf{X}_V^T$, arranged in decreasing order $\left(\lambda_{V_1} \geq \ldots \lambda_{V_{L_{sum}}} \geq 0 \right)$ and $U_{V_1}, \ldots, U_{V_{L_{sum}}}$, the corresponding eigenvectors, where $L_{sum} = L_1 + L_2$. Note also that the structure of the matrix $\mathbf{X}_V \mathbf{X}_V^T$ is as follows:

$$\mathbf{X}_V \mathbf{X}_V^T = \begin{bmatrix} \mathbf{X}^{(1)} \mathbf{X}^{(1)T} & \mathbf{X}^{(1)} \mathbf{X}^{(2)T} \\ \mathbf{X}^{(2)} \mathbf{X}^{(1)T} & \mathbf{X}^{(2)} \mathbf{X}^{(2)T} \end{bmatrix}. \tag{4.5}$$

The structure of the matrix $\mathbf{X}_V \mathbf{X}_V^T$ is similar to the variance-covariance matrix in the classical multivariate statistical analysis literature. The matrix $\mathbf{X}^{(i)} \mathbf{X}^{(i)T}$, which is used in SSA, for the series $Y_{N_i}^{(i)}$, appears along the main diagonal and the products of two Hankel matrices $\mathbf{X}^{(i)} \mathbf{X}^{(j)T}$ $(i \neq j)$, which are related to the series Y_1 and Y_2, appears in the off-diagonal. The SVD of $\mathbf{X}_V$ can be written as $\mathbf{X}_V = \mathbf{X}_{V_1} + \cdots + \mathbf{X}_{V_{L_{sum}}}$, where $\mathbf{X}_{V_i} = \sqrt{\lambda_i} U_{V_i} V_{V_i}^T$ and $V_{V_i} = \mathbf{X}_V^T U_{V_i} / \sqrt{\lambda_{V_i}}$

($\mathbf{X}_{V_i} = 0$ if $\lambda_{V_i} = 0$). In what follows we briefly outline the VMSSA forecasting algorithms. Given that this is a new theoretical development, it is pertinnt to recall the VMSSA forecasting algorithms and in doing so Hassani and Mahmoudvand (2013) is mainly followed.

4.2.1 VMSSA Recurrent Forecasting Algorithm (VMSSA-R)

Let us have two series with different length $Y_{N_i}^{(i)} = (y_1^{(i)}, \ldots, y_{N_i}^{(i)})$ and corresponding window length L_i, $1 < L_i < N_i, i = 1, 2$. The VMSSA-R forecasting algorithm for the h-step ahead forecast is as follows.

1. For a fixed value of K, construct the trajectory matrix $\mathbf{X}^{(i)} = [X_1^{(i)}, \ldots, X_K^{(i)}] = (x_{mn})_{m,n=1}^{L_i,K}$ for each single series $Y_{N_1}^{(1)}$, and $Y_{N_2}^{(2)}$ separately.

2. Construct the block trajectory matrix $\mathbf{X}_V$ as follows:

$$\mathbf{X}_V = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}. \tag{4.6}$$

3. Denote $\lambda_{V_1} \geq \ldots \geq \lambda_{V_{L_{sum}}} \geq 0$ are the eigenvalues of the $\mathbf{X}_V \mathbf{X}_V^T$, where $L_{sum} = L_1 + L_2$.

4. Let $\mathbf{U}_{V_j} = (U_j^{(1)}, U_j^{(2)})^T$ be the jth eigenvector of the $\mathbf{X}_V \mathbf{X}_V^T$, where $U_j^{(i)}$ with length L_i corresponds to the series $Y_{N_i}^{(i)}$ ($i = 1, 2$).

5. Consider $\widehat{\mathbf{X}}_V = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} U_{V_i} U_{V_i}^T \mathbf{X}_V$ as the reconstructed matrix achieved from r eigentriples:

$$\widehat{\mathbf{X}}_V = \begin{bmatrix} \widehat{\mathbf{X}}^{(1)} \\ \widehat{\mathbf{X}}^{(2)} \end{bmatrix}. \tag{4.7}$$

6. Consider matrix $\widetilde{\mathbf{X}}^{(i)} = \mathscr{H}\widehat{\mathbf{X}}^{(i)}$ $(i = 1, 2)$ as the result of the Hankelization procedure of the matrix $\widehat{\mathbf{X}}^{(i)}$ obtained from the previous step, where $\mathscr{H}$ is a Hankel operator.

7. Assume $U_j^{(i)\nabla}$ denotes the vector of the first $L_i - 1$ components of the vector $U_j^{(i)}$ and $\pi_j^{(i)}$ is the last component of the vector $U_j^{(i)}$ $(i = 1, 2)$.

8. Select the number of r eigentriples for the reconstruction stage that can also be used for forecasting purpose.

9. Define matrix $\mathbf{U}^{\nabla(1,2)} = \left(U_1^{\nabla(1,2)}, \ldots, U_r^{\nabla(1,2)} \right)$, where $U_j^{\nabla(1,2)}$ is as follows:

$$U_j^{\nabla(1,2)} = \left[\begin{array}{c} U_j^{(1)\nabla} \\ U_j^{(2)\nabla} \end{array} \right]. \tag{4.8}$$

10. Define matrix $\mathbf{W}$ as follows:

$$\mathbf{W} = \left[\begin{array}{cccc} \pi_1^{(1)} & \pi_2^{(1)} & \cdots & \pi_r^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \cdots & \pi_r^{(2)} \end{array} \right]. \tag{4.9}$$

11. If the matrix $\left(\mathbf{I}_{2\times 2} - \mathbf{W}\mathbf{W}^T \right)^{-1}$ exists and $r \leq L_{sum} - 2$, then the h-step ahead VMSSA forecasts exist and is achieved by the following formula:

$$\left[\hat{y}_{j_1}^{(1)}, \hat{y}_{j_2}^{(2)} \right]^T = \begin{cases} \left[\tilde{y}_{j_1}^{(1)}, \tilde{y}_{j_2}^{(2)} \right], & j_i = 1, \ldots, N_i \\ \\ \left(\mathbf{I}_{2\times 2} - \mathbf{W}\mathbf{W}^T \right)^{-1} \mathbf{W}\mathbf{U}^{\nabla 2^T}\mathbf{Z}_h, & j_i = N_i + 1, \ldots, N_i + h, \end{cases} \tag{4.10}$$

where, $\mathbf{Z}_h = \left[Z_h^{(1)}, Z_h^{(2)} \right]^T$ and $Z_h^{(i)} = \left[\hat{y}_{N_i-L_i+h+1}^{(i)}, \ldots, \hat{y}_{N_i+h-1}^{(i)} \right]$ $(i = 1, 2)$. It should be noted that equation (4.10) indicates that the h-step ahead forecasts of the refined series $\hat{Y}_{N_i}^{(i)}$ are obtained by a multi dimensional linear

recurrent formula (LRF). For the univariate case, there is only one dimensional LRF.

4.2.2   VMSSA Vector Forecasting Algorithm (VMSSA-V)

Let us have items (1)-(10) of VMSSA-R. Consider the matrix:

$$\Pi = \mathbf{U}^{\triangledown}\mathbf{U}^{\triangledown T} + \mathscr{R}\left(\mathbf{I}_{2\times 2} - \mathbf{W}\mathbf{W}^{T}\right)\mathscr{R}^{T}, \tag{4.11}$$

where, $\mathscr{R} = \mathbf{U}^{\triangledown}\mathbf{W}^{T}\left(\mathbf{I}_{2\times 2} - \mathbf{W}\mathbf{W}^{T}\right)^{-1}$.

Let $\Pi = \left(\Pi^{(1)}, \Pi^{(2)}\right)^{T}$ and $\mathscr{R} = \left(\mathscr{R}^{(1)}, \mathscr{R}^{(2)}\right)^{T}$, where $\Pi^{(i)}$ with dimension $(L_i - 1) \times (L_{sum} - 2)$ and $\mathscr{R}^{(i)}$ $(i = 1, 2)$ with length $L_{sum} - 2$ correspond to the series $Y_{N_i}^{(i)}$. Then, Theorem 1 in Hassani and Mahmoudvand (2013) indicates that the linear projection $\mathscr{P}^{(v)} : \mathfrak{L}_r \mapsto \mathbb{R}^{L_{sum}-2}$ by the following formula provides the continuation vectors for the multivariate V-forecasting.

$$\mathscr{P}^{(v)}Y = \begin{bmatrix} \Pi^{(1)}Y_{\triangle} \\ \mathscr{R}^{(1)T}Y_{\triangle} \\ \Pi^{(2)}Y_{\triangle} \\ \mathscr{R}^{(2)T}Y_{\triangle} \end{bmatrix}, Y \in \mathfrak{L}_r, \tag{4.12}$$

where, $Y_{\triangle}^{T} = \left(Y_{\triangle}^{(1)}, Y_{\triangle}^{(2)}\right)$ such that $Y_{\triangle}^{(i)}(i = 1, 2)$ denotes the last $L_i - 1$ entities of Y_i with length L_i. Using above notations, the following algorithm is proposed for calculating the VMSSA-V forecasts.

1. Define vectors Z_i as follows:

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \dots, k \\ \mathscr{P}^{(v)}Z_{i-1} & \text{for } i = k+1, \dots, k+h+L_{\max}-1, \end{cases} \tag{4.13}$$

where, $L_{\max} = \max\{L_1, L_2\}$.

2. Constructing the matrix $\mathbf{Z} = [Z_1 : \ldots : Z_{K+h+L_{\max}-1}]$ and making its han-kelization. Using this calculation we obtain $\hat{y}_1^{(i)}, \ldots, \hat{y}_{N+h+L_{\max}}^{(i)}$ $(i = 1, 2)$.

3. The numbers $\hat{y}_{N_i+1}^{(i)}, \ldots, \hat{y}_{N_i+h}^{(i)}$ $(i = 1, 2)$ form the h step ahead VMSSA-V forecasts.

4.3 Applications

This section considers applications of the proposed theory under various scenarios which include official forecasts, professional forecasts and forecasts from other time series models. In the real world, publishers of official forecasts are usually interested in providing predictions for the coming year (i.e. 12 steps ahead for monthly data and 4 steps ahead for quarterly data). In line with this, considered here are applications which provide out-of-sample forecasts for the next year. For example, if we are dealing with monthly data, the last 12 observations for which 12 forecasted data are available are set aside as the out-of-sample data and the remainder is used for training and testing the forecasting models which are used for comparison purposes. Where 12 observations are forecasted, this means the first forecasted data point is the $h = 1$ step ahead forecast, the second forecasted data point is the $h = 2$ steps ahead forecast and so on up until the final forecasted data point which will represent the $h = 12$ steps ahead forecast or the 12 months ahead value of a given variable. All applications consider the RMSE as the loss function and all outcomes are evaluated for statistical significance using the Diebold-Mariano (DM) test in Diebold and Mariano (1995) and the Kolmogorov-Smirnov Predictive Accuracy (KSPA) test in Chapter 3. It should be noted that as a result of a small number of observations for evaluating forecast accuracy, it is likely that statistical tests will experience issues with picking up significant outcomes.

4.3.1   Using Forecasts from Other Forecasting Models as More Information

Considered in the first scenario is the use of forecasts from a variety of other time series analysis models such as ARIMA, Exponential Smoothing and Holt-Winters for improving the accuracy of these forecasts further via VMSSA. This is important as especially in government organizations, methods such as ARIMA and Holt-Winters are widely accepted and continue to be used owing to traditions and familiarity with such models. Figure 1 plots the time series used here as examples. Each time series has been obtained via Datamarket[1].

These monthly time series include the popular U.S. Accidental deaths time series (monthly data, 78 observations, January 1973-June 1979), milk production (monthly data, 168 observations, January 1962-December 1975), number of city births in New York over time (monthly data, 168 observations, January 1946-December 1959) and residential electricity usage in Iowa, U.S. (monthly data, 106 observations, January 1971-October 1979). It is clear via Figure 4.1 that the series chosen via Datamarket include those which captures stationarity, non-stationarity, increasing trends, seasonality and structural breaks. In reality we are likely to be faced with such varying time series and it is therefore important to consider such phenomenons as examples. In addition, considered as an example is also an application which seeks to forecast international tourist arrivals into Germany compiled via the Eurostat database (monthly data, 168 observations, January 2000-December 2013). In each case, the last 12 monthly observations are left aside as out-of-sample and the models are trained over the remainder of the observations.

The results from the applications are presented in Table 4.1. The out-of-sample forecasting RMSE's are obtained via ARIMA, HW, ETS, SSA-V, SSA-

---

[1]https://datamarket.com/
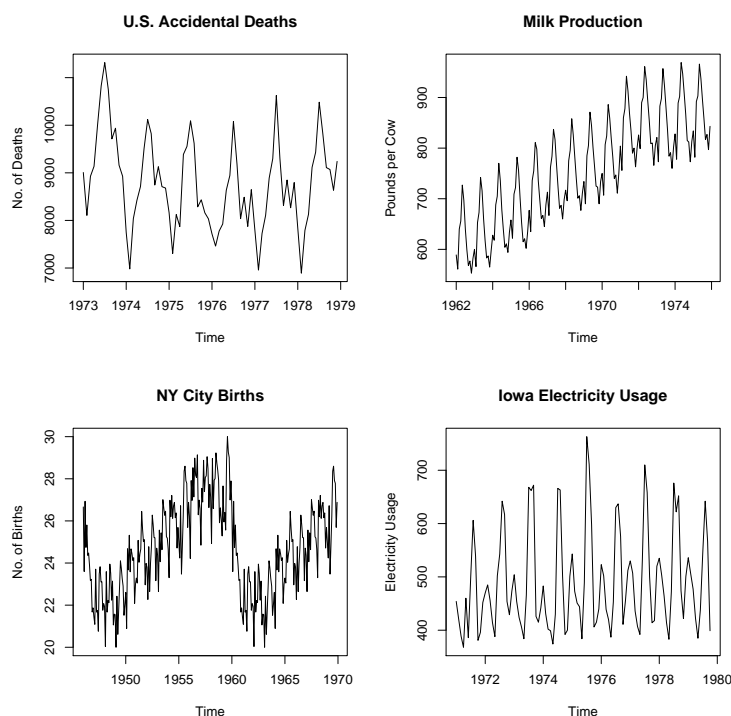
Fig. 4.1 Four time series used as examples from Datamarket.

R, VMSSA-V and VMSSA-R for each data set. Note that when modelling with MSSA, forecasts from the univariate model which reports the lowest in-sample forecasting RMSE for the training data is selected as more information in the MSSA model to obtain out-of-sample forecasts.

The U.S. Accidental deaths series has been widely adopted in time series literature, see for example Hassani (2007) and Brockwell and Davis (2002). For the death series, ARIMA provided the lowest in-sample forecasting RMSE and therefore the out-of-sample forecasts from ARIMA were considered as additional information for the MSSA model. It is important to evaluate whether the newly proposed approach can result in forecasts which not only outperform the accuracy of the initial forecast, but also forecasts from SSA. Based on the RMSE criterion, it is evident that VMSSA can provide forecasts with the lowest RMSE in comparison to all other models for this series, and VMSSA-V in

particular reports the lowest RMSE. However, these forecasting differences can be attributed to chance occurrences. In order to evaluate if the VMSSA forecast is significantly better, all outcomes are tested for statistical significance with the results being reported in Table 4.2 along with the RRMSE.

Based on the RRMSE, the VMSSA-V forecasts are 6% better than ARIMA, 28% better than HW, 8% better than ETS, 58% better than SSA-V forecasts. Likewise, VMSSA-R forecasts are 2% better than ARIMA, 24% better than HW, 3% better than ETS and 48% better than SSA-R. In this case, based on both DM and KSPA tests we find evidence of statistically significant differences between the forecasts of VMSSA-V and SSA-V, and VMSSA-R and SSA-R at the 10% significance level. However we do not find similar evidence in relation to the other models. Yet, the fact that VMSSA forecasts are significantly better than the SSA forecasts indicate that the proposed approach is viable.

Table 4.1 RMSE for forecasting last year of each data set.

Series	ARIMA	HW	ETS	SSA-V	SSA-R	VMSSA-V	VMSSA-R
Death	332	432	338	736	624	312	327
Milk Prod.	14.10	14.80	8.63	19.50	13.70	7.28	7.69
NY Births	0.91	1.06	1.13	1.38	1.46	0.85	0.88
Elec. Use	51.6	78.10	39.90	57.30	53.70	38.73	36.40
Tourism	58251	80504	82217	55334	42089	46211	40010

Note: Forecasts from the univariate model providing the lowest in-sample forecasting RMSE is used as additional information for the MSSA model.

Table 4.2 RRMSE for forecasting last year of each data set.

Series	$\frac{VMSSA-V}{ARIMA}$	$\frac{VMSSA-R}{ARIMA}$	$\frac{VMSSA-V}{HW}$	$\frac{VMSSA-R}{HW}$	$\frac{VMSSA-V}{ETS}$	$\frac{VMSSA-R}{ETS}$	$\frac{VMSSA-V}{SSA-V}$	$\frac{VMSSA-R}{SSA-R}$
Death	0.94	0.98	0.72	0.76	0.92	0.97	$0.42^{*,\dagger}$	$0.52^{*,\dagger}$
Milk Prod.	$0.52^{*,\dagger}$	0.55^{*}	$0.49^{*,\dagger}$	$0.52^{*,\dagger}$	0.84	0.89	$0.37^{*,\dagger}$	0.56^{*}
NY Births	0.93	0.97	0.80	0.83	0.75	0.76	0.62	0.60
Elec. Use	0.75	0.71	$0.50^{*,\dagger}$	$0.47^{*,\dagger}$	0.97	$0.91^{\dagger}$	$0.68^{*,\dagger}$	$0.68^{*,\dagger}$
Tourism	0.79	0.69	$0.57^{*,\dagger}$	$0.50^{*,\dagger}$	$0.56^{*,\dagger}$	$0.49^{*,\dagger}$	$0.84^{\dagger}$	0.95

Note: * indicates a statistically significant difference between the two forecasts based on the modified Diebold-Mariano test at $p = 0.10$. † indicates a statistically significant difference between the two forecasts based on the KSPA test at $p = 0.10$.

Considered next is the monthly milk production series. In this case, ETS forecasts were found to be best in-sample and is therefore considered as the additional information for the VMSSA model. Once again, based on the RMSE

in Table 4.1, it is clear that the VMSSA forecasts can outperform the rest of the models considered here. These performances convert such that VMSSA-V forecasts are 48% better than ARIMA, 51% better than HW, 16% better than ETS, and 63% better than SSA-V forecasts (Table 4.2). Likewise, VMSSA-R forecasts are 45% better than ARIMA, 48% better than HW, 11% better than ETS, and 44% better than SSA-R forecasts (Table 4.2). Interestingly, in relation to the previous application, there are a higher number of statistically significant outcomes in this case. VMSSA forecasts via the proposed approach are significantly better than ARIMA, ETS and SSA-V and SSA-R forecasts.

The third application in this sub section considers monthly city births in New York. In this instance ARIMA provided the best in-sample forecast and was therefore selected as the model which will provide more information for the VMSSA process. Table 4.1 shows that VMSSA once again outperforms all models based on the RMSE, and that VMSSA-V records the lowest RMSE. The RRMSE values in Table 4.2 indicates that VMSSA-V forecasts are 7%, 20%, 25% and 38% better than ARIMA, HW, ETS and SSA-V forecasts respectively whilst VMSSA-R forecasts are 3%, 7%, 24% and 40% better than ARIMA, HW, ETS and SSA-V forecasts respectively. Regardless of the gains suggested via the RRMSE criterion there is no sufficient evidence of statistically significant differences between the VMSSA and competing forecasts in this case. Given the comparatively large gains reported here, the inability of the statistical tests at picking up significant differences could be a result of small sample sizes.

The fourth application relating to the use of forecasts from other models as more information looks at monthly average residential electricity usage in Iowa. ETS provided the best in-sample forecast for this series and therefore its out-of-sample forecast was considered as more information in the VMSSA framework. As reported in Table Table 4.1, once again the VMSSA models outperform the rest based on the RMSE criterion. The RRMSE indicates that

VMSSA-V forecast reports gains of 25%, 50%, 3% and 32% in relation to the forecasts from ARIMA, HW, ETS and SSA-V respectively. At the same time, VMSSA-R forecast reports gains of 29%, 53%, 9%, 32% in relation to the forecasts from ARIMA, HW, ETS and SSA-V respectively. The tests for statistical significance indicates there exists significant differences between VMSSA and HW forecasts and VMSSA and SSA forecast. In addition, there is a statistically significant difference between the VMSSA-R and ETS forecast.

The final application here looks at the monthly international tourist arrivals into Germany. The univariate SSA-R forecast was seen providing the most accurate in-sample forecast for these series based on the lowest RMSE and therefore forecasts generated by this model over the out-of-sample period was considered as more information. The RMSE results reported in Table 4.1 shows that in terms of the univariate models, SSA-R reports the forecast with the lowest error. However, an application of the new approach proposed in this chapter results in forecasts by VMSSA-R which outperforms the rest of the models. Table 4.2 indicates that the VMSSA-R forecasts are 31%, 50%, 51%, and 5% better than ARIMA, HW, ETS and SSA-R forecasts respectively with statistically significant differences reported between VMSSA-R forecasts and those of HW and ETS.

4.3.2   Using Official Forecasts as More Information

Having considered the use of forecasts calculated from other models as auxiliary information for the VMSSA process, this sub section looks at exploiting the forecastability of official forecasts. As such, these official or professional forecasts which are calculated using complex multivariate models in most instances are considered as more information. The applications consider as official forecasts (OF), those obtained via the U.S. Energy Information Administration

(EIA)[2] for a variety of variables and a professional forecast (PF) for inflation by a group of non-financial service providers who include manufacturers, universities, forecasting firms, investment advisors, pure research firms and consulting firms. The EIA time series are shown in Figure 4.2 whereby all data are monthly and the Oil Price, Gas Price, Electricity Sales and Electricity Series each have 81 observations from January 2008-September 2014. The CPI time series for which a professional forecast is available is shown in Figure 4.3 and this series includes quarterly data with 126 observations recorded between Q3 of 1981 and Q4 of 2012. The last year is considered as out-of-sample data and the results from the forecasting exercise are reported in Tables 4.3 and 4.4.
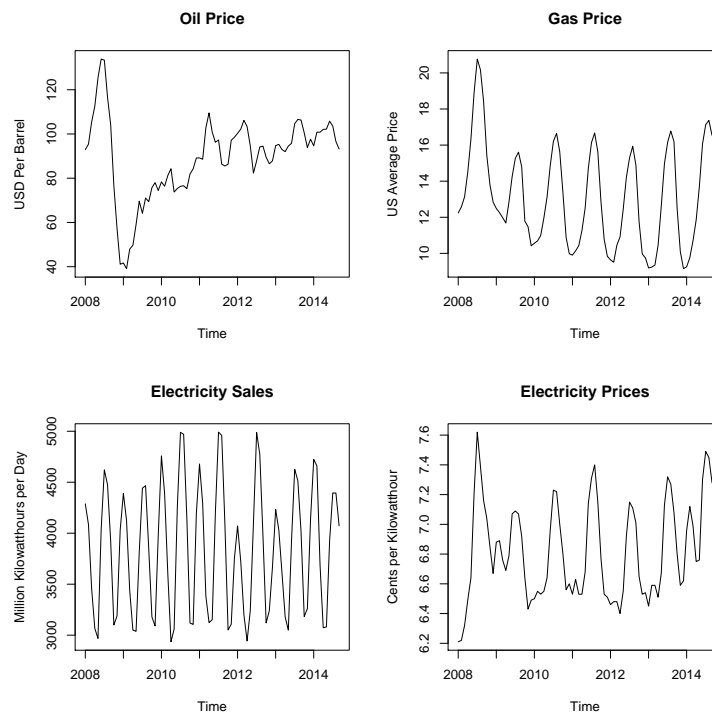
Fig. 4.2 The EIA time series used as examples.

The first application looks at the West Texas Intermediary (WTI) oil price series. The RMSE results in Table 4.3 shows that VMSSA outperforms the EIA

---

[2]http://www.eia.gov/forecasts/steo/outlook.cfm
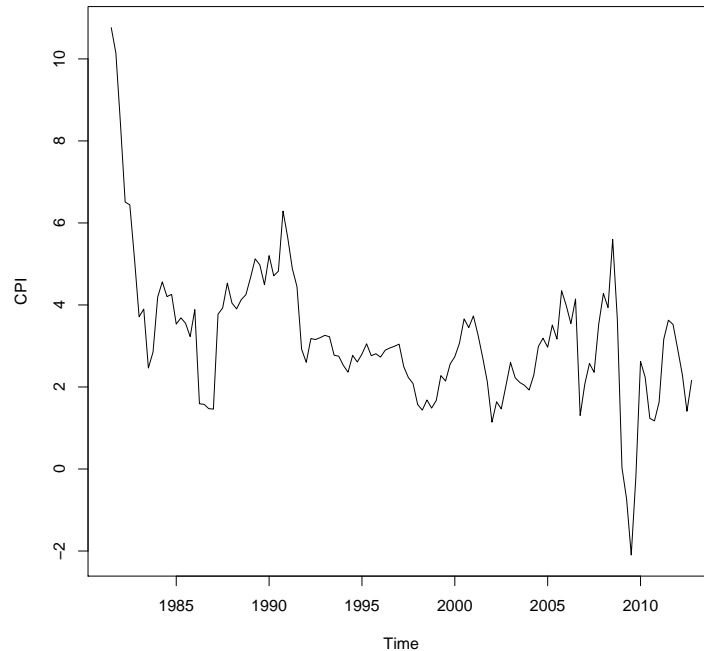
Fig. 4.3 Quarterly consumer price index data.

official forecast and also the SSA forecasts along with the VMSSA-V model reporting the lowest RMSE. In terms of the RRMSE criterion, as reported in Table 4.4, the VMSSA-V forecasts are 2% better than the official forecasts, and 22% better than the SSA-V forecasts. Likewise, the VMSSA-R forecasts are 1% better than the official forecasts, and 57% better than the SSA-R forecasts. However, when tested for statistically significant differences between the forecasts evidence was found only for significant differences between the VMSSA-R and SSA-V forecasts.

The next application considers average residential natural gas prices in the United States. Again, based on the RMSE values the VMSSA forecasts outperform the official EIA forecast and the SSA forecasts and VMSSA-V reports the lowest RMSE (Table 4.3. The RRMSE values in this case show very minor gains whereby VMSSA-V is 6% better than the official forecast and 9% bet-

ter than the SSA-V forecast, whereas the VMSSA-R forecast is 5% better than the official forecast and 8% better than the SSA-R forecast. Here there is no evidence of statistically significant differences between any of the forecasts.

Table 4.3 RMSE when using official forecasts for forecasting last year of each data set.

Series	OF	SSA-V	SSA-R	VMSSA-V	VMSSA-R
EIA					
WTIPUUS	4.34	5.42	10.01	4.25	4.32
NGRCUUS	0.87	0.90	0.90	0.82	0.83
EXRCP.US	253.46	392.69	306.44	253.95	248.53
ESICU.US	0.24	0.31	0.31	0.23	0.23
PF					
CPI	0.53	1.17	2.40	0.31	0.40

Note: OF: Official forecast. WTIPUUS - West Texas Intermediate Spot Average Crude Oil Price. NGRCUUS - Average Residential Natural Gas Price. EXRCP.US - Residential Sector Total Electricity Sales. ESICU.US - Industrial Sector Average Regional Electricity Prices. PF: Professional forecast from group of non-financial service providers. CPI - Consumer Price Index.

The third application which considers an official forecast looks at data on total electricity sales in the U.S. residential sector. The RMSE results in Table 4.3 shows that VMSSA-R can provide the forecast with the lowest error whilst the VMSSA-V forecast is on par with the official forecast. The RRMSE values in Table 4.4 indicates that VMSSA-V forecast is 35% better than the SSA-V forecast whilst the VMSSA-R forecast is 2% better than the official forecast and 19% better than the SSA-R forecast. All outcomes are once again tested for statistical significance, but there is no evidence at the 10% significance level in this case.

The fourth application considers modelling the U.S. industrial sector average regional electricity prices. The RMSE values in Table 4.3 shows that both MSSA models are outperforming the official forecast and the SSA models in this case. The RRMSE criterion as per Table 4.4 indicates that both MSSA forecasts are 4% better than the official forecast and 26% better than the SSA-R forecast. In this case there is evidence of the newly proposed VMSSA approach

outperforming SSA forecasts with statistically significant results.

Table 4.4 RRMSE when using official forecasts for forecasting last year of each data set.

Series	$\frac{VMSSA-V}{OF}$	$\frac{VMSSA-R}{OF}$	$\frac{VMSSA-V}{SSA-V}$	$\frac{VMSSA-R}{SSA-R}$
EIA				
WTIPUUS	0.98	0.99	0.78	$0.43^{*,\dagger}$
NGRCUUS	0.94	0.95	0.91	0.92
EXRCP.US	1.00	0.98	0.65	0.81
ESICU.US	0.96	0.96	0.74^{*}	$0.74^{*,\dagger}$
PF				
CPI	0.58	0.75	0.26	0.17

Note: * indicates a statistically significant difference between the two forecasts based on the modified Diebold-Mariano test at $p = 0.10$. † indicates a statistically significant difference between the two forecasts based on the KSPA test at $p = 0.10$.

The final application considers forecasting the last four quarters of the quarterly consumer price index growth rate series. The professional forecast is used as more information and the resulting RMSE is reported in Table 4.3. Given that there are only four out-of-sample observations it isn't realistic to expect statistically significant differences between the forecasts in this case. However, the RRMSE results in Table 4.4 can provide a reasonable indication of the comparative performance. The RMSE shows that both VMSSA models outperform not only SSA but also the professional forecast with VMSSA-V reporting the lowest RMSE. The RRMSE criterion shows that the VMSSA-V forecast is 42% better than the professional forecast and 74% better than the SSA-V forecast. Likewise, the VMSSA-R forecast is 25% better than the professional forecast and 83% better than the SSA-R forecast. Figure 4.4 provides a graphical representation of the out-of-sample forecasts. It is evident that the VMSSA-V forecast is the only one which remains comparatively aligned with the actual inflation values.

Inflation Forecasts
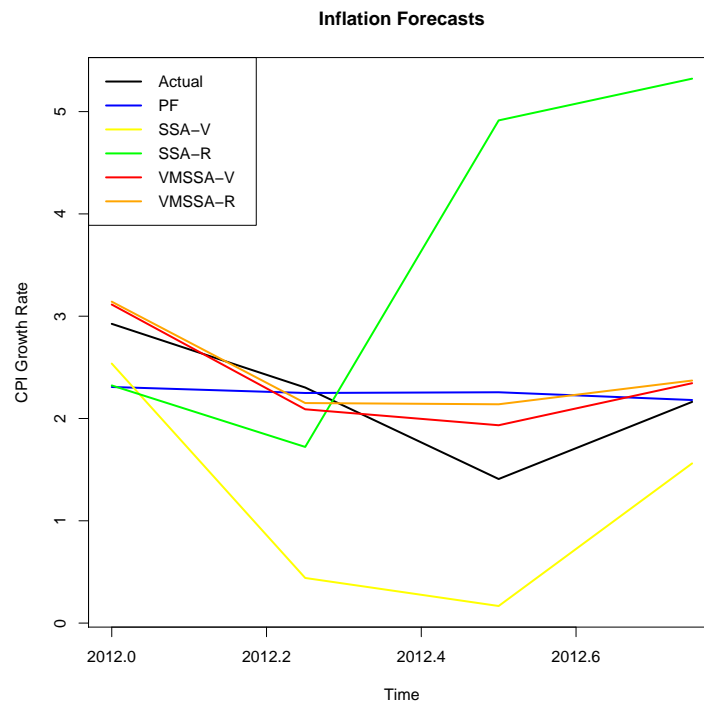
Fig. 4.4 Out-of-sample forecasts for the last four quarters of the CPI.

4.4   Discussion

This chapter begins with the aim of introducing a theoretical framework for exploiting the forecastability of forecasts. That is, once a forecast is generated by official or professional forecasters, is there any possibility of exploiting the information contained with the given forecast for generating a new and more accurate forecast? The idea for exploiting the forecastability of forecasts was derived from the methodology underlying VMSSA (Hassani and Mahmoudvand, 2013) which is a technique that enables modelling multiple time series with different series lengths.

The proposed methodology seeks to exploit data with a time lag into the future and couples this information with historical data pertaining to the same variable in order to generate a new and improved forecast. The only condi-

tion is that the forecast has to have some level of good accuracy as otherwise there would not be any useful auxiliary information that can be extracted from the forecast. Given the proposed theory, this chapter applies it to several real world applications. The results indicate that VMSSA forecasts which exploits the proposed theory are able to outperform its univariate counterpart, SSA in all instances (with statistically significant results in some cases). Moreover, there has always been at least one VMSSA model which can outperform the official, professional or other forecasts in all cases based on the RMSE criterion. The low number of out-of-sample forecast available for comparison purposes makes it an arduous task for the statistical tests to pick up significant differences. However, the RRMSE criterion is able to show that in certain cases the VMSSA models report gains of well over 20% in relation to a competing forecast.

The introductory nature of this theoretical concept opens up a new research avenue with specific interests for the future discussed in the final chapter. However, the initial findings not only introduces a novel theoretical approach for exploiting the forecastability of forecasts, but also shows that it is indeed worthwhile to research in-depth into this concept so as to develop more efficient VMSSA models which will be of utmost importance to forecasters across the globe.

Chapter 5

Automated & Optimized Singular Spectrum Analysis

This chapter focuses on introducing the automated and optimized SSA and MSSA forecasting algorithms. The chapter begins with an introduction which considers the reasons for, and the importance of automating and optimizing the SSA and MSSA processes which is then followed by a concise literature review which evaluates the historical approaches to parameter selection in SSA and MSSA whilst indicating how the proposed method fills an existing research gap. The algorithms are presented next, and an application of the automated SSA process to forecasting a real data set is presented as empirical evidence.

5.1   Introduction

In the 21st century, Econometricians are no longer the only individuals who are interested in time series analysis and forecasting as both large and small scale firms now seek to exploit time series analysing and forecasting methods to enable better decision making and risk management. The increasing availability of large data sets and access to time series information on the world wide web has

further added to the interest in exploiting this widely available information for improving managerial decisions. However, this increasing interest by the masses towards the application of time series analysis techniques is greatly restricted by the nature of the time series techniques themselves. This is because, with the exception of the Random Walk, almost every other time series technique has a complex econometric framework underlying its understanding, usage and performance which in turn restricts the ability of those who are not fortunate enough to comprehend such complexities from exploiting these methods.

In 2008, researchers from Monash University in Australia having understood the need for large scale forecasting and the lack of trained personnel in the field of time series analysis and forecasting techniques even when a small number of forecasts are required, introduced the 'forecast package' for R which in brief enables automatic time series forecasting (Hyndman and Khandakar, 2008). This meant that individuals no longer need to understand the complex economic theory underlying methods such as ARIMA, Exponential Smoothing, and Neural Networks (among other techniques introduced through this package) in order to generate forecasts using these methods. The concept was well received and the authors went a step further to complement the 'forecast package' with an online text book on forecasting using these methods, see Hyndman and Athanasopoulos (2012).

The success of the Singular Spectrum Analysis technique depends solely on the correct specification of its two parameters, the Window Length L and the number of eigenvalues r (Hassani and Mahmoudvand, 2013; Hassani et al., 2011, 2012b), which are used for decomposing a time series and then reconstructing a less noisy time series respectively. Over the years, a variety of mathematically complex, time consuming and labour intensive approaches which require detailed knowledge on the theory underlying SSA have been proposed and developed for the selection of SSA and MSSA parameters. However, the highly

labour intensive and complex nature of such approaches have not only discouraged the application of this method by those not conversant with the underlying theory, but also limited SSA and MSSA to offline applications. As such there remains a need for automating SSA and MSSA. At this juncture it is pertinent to note some key initial attempts at automating SSA. Firstly, Vautard et al. (1992) sought to automate trend and periodicity extraction in SSA. Several years later, Alonso et al. (2004) presented automated denoising for SSA in the case of big signal to noise ratios. This was followed by Alexandrov (2009) who presented a method for automating trend extraction using SSA. Whilst the previous automation attempts have lead to positive outcomes in terms of enabling users to extract trend and harmonic components with ease, these methods are not aimed at enabling the best possible forecast from SSA or MSSA.

Accordingly, through this research, an automated and optimized algorithm is presented for forecasting with SSA and MSSA. The key point being that this algorithm is optimized by minimizing a loss function which enables the users to automatically determine the optimal SSA or MSSA parameters for obtaining the best possible forecast without the need for an extensive or in-depth knowledge into the complex theory underlying SSA and MSSA.

5.2   Parameter Selection in SSA & MSSA

The selection of the window length L depends on the structure of the data, the purpose of the analysis and the forecasting horizon (Hassani et al., 2009; Hassani and Mahmoudvand, 2013). Following some discussion, Elsner and Tsonis (1996) notes that selecting $L = T/4$ is common practice. However, theoretical results thereafter suggest that the window length L should be large enough but not greater than $N/2$ (Ghodsi et al., 2009; Golyandina et al., 2001; Hassani, 2007). The selection of L is both crucial and problematic as when L is too large

there are too few observations left to choose the covariance matrix of the L variables (Hassani et al., 2011) and this is known to make the forecasting results inaccurate (Hassani et al., 2012b). Moreover, setting L too large could lead to some parts of the noise mixing up with the signal whilst choosing L too small opens up the risk of losing some parts of the signal to the noise (Golyandina et al., 2001). Hassani (2007) notes that if the time series has a periodic component with an integer period, then it is advisable to select L proportional to that period as this enables better separability of the periodic components.

Golyandina (2010) recommends setting L close to half of the time series length to achieve optimal signal-noise separation based on evidence from a simulation study. However, Khan and Poskitt (2013a) provides evidence which shows that the Golyandina (2010) claim does not hold universally. Hassani et al. (2011) and Hassani et al. (2012b) suggested considering the selection of L based on the concept of separability between signal and noise. The authors show that by setting $L = \left[\frac{T+1}{2}\right]$ where T is the length of the series, one is able to attain the minimum value for the weighted correlation (w-correlation) statistic which is a natural measure of the similarity between two series (Hassani et al., 2011). Through the work of Khan and Poskitt (2013b) it is suggested that setting L much shorter than the upper bound $N/2$ can result in better SSA forecasts. In other words, their recommendation is that $L \leq\leq\leq\leq N/2$ and $L = (\log N)^c$, where $c > \log(2)/\log\log(N)$ (Khan and Poskitt, 2013b). In summary, there is no one rule for selecting L, and instead it largely depends on the structure of the data, purpose of analysis and the forecasting horizon of interest.

The selection of the correct number of eigenvalues r is equally important in the overall SSA process as it has a direct effect on the reconstruction in SSA. As Hassani and Mahmoudvand (2013) notes, if r is chosen to be greater than exactly what it should be, then we increase the noise in the reconstructed series whereas choosing r to be smaller than the exact requirement results in ignor-

ing some parts of the signal which ought to be included in the reconstruction. Literature shows that there are various approaches to select r. Hassani (2007) suggests analysing the scree plot and pairwise scatter plots. However, as Khan and Poskitt (2013b) points out there is no defined statistical decision rules when using these approaches and so the modelling procedure is left to be a highly subjective assessment. Accordingly, the selection of r in SSA continues to remain an open problem. Parameter selection is even more complex in the case of MSSA. As Hassani and Mahmoudvand (2013) states, the similarity and orthogonality among series, the use of a block trajectory Hankel matrix as opposed to one Hankel matrix makes the selection of L a difficult task whereas the selection of r is made difficult by the fact that each eigenvalue contains information of all time series considered in the multivariate analysis.

5.3   New Forecasting Algorithms for SSA and MSSA

Presented in this section are new forecasting algorithm for SSA and MSSA which enables the automatic selection of L and r for obtaining the optimal SSA or MSSA forecast for a given data set. The algorithms are optimized via the minimisation of a loss function and is the first step towards enabling the use of SSA and MSSA for online applications. Moreover, the automated nature of the algorithms enable users who are not conversant with the complex theory underlying SSA and MSSA to be able to exploit these techniques for their work.

5.3.1   Automated & Optimized SSA Forecasting Algorithm

Shown below is the automated and optimized SSA forecasting algorithm.

1. Consider a real-valued nonzero time series $Y_N = (y_1, \ldots, y_N)$ of length N.

2. Divide the time series into two parts; $\frac{2}{3}^{rd}$ of observations for model training and testing, and $\frac{1}{3}^{rd}$ for validating the selected model.

3. Use the training data to construct the trajectory matrix $\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = [X_1, ..., X_K]$, where $X_j = (y_j, ..., y_{L+j-1})^T$ and $K = N - L + 1$. Initially, we begin with $L = 2$ ($2 \leq L \leq \frac{N}{2}$) and in the process, evaluate all possible values of L for Y_N.

4. Obtain the SVD of $\mathbf{X}$ by calculating $\mathbf{X}\mathbf{X}^T$ for which $\lambda_1, ..., \lambda_L$ denotes the eigenvalues in decreasing order ($\lambda_1 \geq ... \lambda_L \geq 0$) and by $U_1, ..., U_L$ the corresponding eigenvectors. The output of this stage is $\mathbf{X} = \mathbf{X}_1 + ... + \mathbf{X}_L$ where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ and $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$.

5. Evaluate all possible combinations of r ($1 \leq r \leq L - 1$) singular values (step by step) for the selected L and split the elementary matrices $\mathbf{X}_i$ ($i = 1, ..., L$) into several groups and sum the matrices within each group.

6. Perform diagonal averaging to transform the matrix with the selected r singular values into a Hankel matrix which can then be converted into a time series (the steps up to this stage filters the noisy series). The output is a filtered series that can be used for forecasting.

7. Depending on the forecasting approach one wishes to use, select the SSA-R approach or SSA-V approach which are explained below at the end of this algorithm.

8. When forecasting a series Y_N h-steps ahead, the forecast error is minimised by setting $\mathscr{L}(X_{K+h} - \hat{X}_{K+h})$ where the vector $\hat{X}_{K+h}$ contains the h-step ahead forecasts obtained using the corresponding forecasting algorithm as chosen in Step 7.

9. Find the combination of L and r which minimises $\mathscr{L}$ and thus represents the optimal SSA choices of L and r.

10. Finally use the optimal L to decompose the series comprising of the validation set, and then select r singular values for reconstructing the less noisy time series, and use this newly reconstructed series for forecasting the remaining $\frac{1}{3}^{rd}$ observations.

Recurrent SSA (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 $U^\nabla \in \mathbf{R}^{L-1}$ 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

$$y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, N \\ \sum_{j=1}^{L-1} \alpha_j y_{i-j} & \text{for } i = N+1, \ldots, N+h \end{cases} \tag{5.1}$$

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:

$$A = \frac{1}{1-v^2} \sum_{i=1}^{r} \pi_i U_i^\nabla. \tag{5.2}$$

Vector SSA (SSA-V)

Consider the following matrix

$$\Pi = \mathbf{V}^\nabla (\mathbf{V}^\nabla)^T + (1-v^2)AA^T \tag{5.3}$$

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

$$\theta^{(v)} : \mathfrak{L}_r \mapsto \mathbf{R}^L \tag{5.4}$$

where

$$\theta^{(v)} U = \begin{pmatrix} \Pi U^{\triangledown} \\ A^T U^{\triangledown} \end{pmatrix}. \tag{5.5}$$

Define vector Z_i as follows:

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, K \\ \theta^{(v)} Z_{i-1} & \text{for } i = K+1, \ldots, K+h+L-1 \end{cases} \tag{5.6}$$

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.

5.3.2   Automated & Optimized MSSA Forecasting Algorithms

In what follows, the automated and optimized HMSSA-R and HMSSA-V forecasting algorithms are presented. In presenting these two algorithms the notations in Hassani and Mahmoudvand (2013) have been relied upon.

HMSSA-R Optimal Forecasting Algorithm

1. Consider M time series with identical series lengths of N_i, such that $Y_{N_i}^{(i)} = (y_1^{(i)}, \ldots, y_{N_i}^{(i)})$ $(i = 1, \ldots, M)$.

2. Split each time series into three parts leaving $\frac{2}{3}^{rd}$ for model training and testing, and $\frac{1}{3}^{rd}$ for validation.

3. Beginning with a fixed value of $L = 2$ ($2 \leq L \leq \frac{N}{2}$) and in the process, evaluating all possible values of L for Y_{N_i}, using the training data construct the trajectory matrix $\mathbf{X}^{(i)} = [X_1^{(i)}, \ldots, X_K^{(i)}] = (x_{mn})_{m,n=1}^{L,K_i}$ for each single series $Y_{N_i}^{(i)}$ ($i = 1, \ldots, M$) separately.

4. Then, construct the block trajectory matrix $\mathbf{X}_H$ as follows:

$$\mathbf{X}_H = \left[\ \mathbf{X}^{(1)} : \ \mathbf{X}^{(2)} : \ \cdots \ : \mathbf{X}^{(M)} \ \right].$$

5. Let vector $U_{H_j} = (u_{1j}, \ldots, u_{Lj})^T$, with length L, be the jth eigenvector of $\mathbf{X}_H \mathbf{X}_H^T$ which represents the SVD.

6. Evaluate all possible combinations of r ($1 \leq r \leq L - 1$) step by step for the selected L and construct $\widehat{\mathbf{X}}_H = \sum_{i=1}^r U_{H_i} U_{H_i}^T \mathbf{X}_H$ as the reconstructed matrix obtained using r eigentriples:

$$\mathbf{X}_H = \left[\ \widehat{\mathbf{X}}^{(1)} : \ \widehat{\mathbf{X}}^{(2)} : \ \cdots \ : \widehat{\mathbf{X}}^{(M)} \ \right].$$

7. Consider matrix $\widetilde{\mathbf{X}}^{(i)} = \mathscr{H}\widehat{\mathbf{X}}^{(i)}$ ($i = 1, \ldots, M$) as the result of the Hankelization procedure of the matrix $\widehat{\mathbf{X}}^{(i)}$ obtained from the previous step for each possible combination of SSA choices.

8. Let $U_{H_j}^{\triangledown}$ denote the vector of the first $L - 1$ coordinates of the eigenvectors U_{H_j}, and π_{H_j} indicate the last coordinate of the eigenvectors U_{H_j} ($j = 1, \ldots, r$).

9. Define $\upsilon^2 = \sum_{j=1}^r \pi_{H_j}^2$.

10. Denote the linear coefficients vector $\mathscr{R}$ as follows:

$$\mathscr{R} = \frac{1}{1 - \upsilon^2} \sum_{j=1}^{r} \pi_{Hj} U_{Hj}^{\triangledown}. \tag{5.7}$$

11. If $\upsilon^2 < 1$, then the h-step ahead HMSSA forecasts exist and is calculated by the following formula:

$$\left[\hat{y}_{j_1}^{(1)}, \ldots, \hat{y}_{j_M}^{(M)}\right]^T = \begin{cases} \left[\tilde{y}_{j_1}^{(1)}, \ldots, \tilde{y}_{j_M}^{(M)}\right], & j_i = 1, \ldots, N_i, \\ \\ \mathscr{R}^T \mathbf{Z}_h, & j_i = N_i + 1, \ldots, N_i + h, \end{cases} \tag{5.8}$$

where, $\mathbf{Z}_h = \left[Z_h^{(1)}, \ldots, Z_h^{(M)}\right]^T$ and $Z_h^{(i)} = \left[\hat{y}_{N_i-L+h+1}^{(i)}, \ldots, \hat{y}_{N_i+h-1}^{(i)}\right]$ $(i = 1, \ldots, M)$.

12. Seek the combination of L and r which minimises a loss function, $\mathscr{L}$ and thus represents the optimal HMSSA-R choices for decomposing and reconstructing in a multivariate framework.

13. Finally use the selected optimal L to decompose the series comprising of the validation set, and then select r singular values for reconstructing the less noisy time series, and use this newly reconstructed series for forecasting the remaining $\frac{1}{3}^{rd}$ observations.

HMSSA-V Optimal Forecasting Algorithm

1. Begin by following the steps in 1-9 of the HMSSA-R optimal forecasting algorithm above.

2. Consider the following matrix

$$\Pi = \mathbf{U}^{\nabla}\mathbf{U}^{\nabla T} + (1 - v^2)RR^T, \qquad (5.9)$$

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

$$\mathscr{P}^{(v)} : \mathfrak{L}_r \mapsto \mathbb{R}^L, \qquad (5.10)$$

where

$$\mathscr{P}^{(v)}Y = \begin{pmatrix} \Pi Y_{\triangle} \\ R^T Y_{\triangle} \end{pmatrix}, \ Y \in \mathfrak{L}_r, \qquad (5.11)$$

and $Y_{\triangle}$ is vector of last $L - 1$ elements of Y.

3. Define vector $Z_j^{(i)}$ $(i = 1, \dots, M)$ as follows:

$$Z_j^{(i)} = \begin{cases} \widetilde{X}_j^{(i)} & \text{for } j = 1, \dots, k_i \\ \mathscr{P}^{(v)}Z_{j-1}^{(i)} & \text{for } j = k_i + 1, \dots, k_i + h + L - 1 \end{cases} \qquad (5.12)$$

where, $\widetilde{X}_j^{(i)}$'s are the reconstructed columns of trajectory matrix of the ith series after grouping and leaving noise components.

4. Now, by constructing matrix $\mathbf{Z}^{(i)} = [Z_1^{(i)}, \dots, Z_{k_i+h+L-1}^{(i)}]$ and performing diagonal averaging we obtain a new series $\hat{y}_1^{(i)}, \dots, \hat{y}_{N_i+h+L-1}^{(i)}$, where $\hat{y}_{N_i+1}^{(i)}, \dots, \hat{y}_{N_i+h}^{(i)}$ provides the h step ahead HMSSA-V forecast for the selected combination of L and r.

5. Finally, follow steps 12-13 in the HMSSA-R optimal forecasting algorithm to find the optimal L and r, and use these to obtain the HMSSA-V forecasts.

5.4   Application

This section considers the application of the automated and optimized SSA algorithm for tourism demand forecasting in order to illustrate its usefulness and validity in practice.

Previous research has highlighted the importance of accurate demand forecasting to the tourism sector. The dependence of tourism on both investment and infrastructure development make a degree of advance planning essential, as many authors have recognised. Well informed investment decisions are vital for efficient resource allocation in both tourism and supporting sectors. The economic downturn and an increased awareness of world economic volatility have strengthened rather than weakened this need to forecast tourist demand accurately.

As discussed in the following sub-section there is an extensive and high profile existing literature on forecasting tourism demand. This literature covers a wide range of different forecasting techniques, applied to a wide range of different countries or locations. The purpose of this application is to add to this literature by introducing a new model for forecasting tourist arrivals and to apply it to inbound U.S. tourist arrivals. Forecasting U.S. tourist arrivals is both a demanding and important task, mainly because these data exhibit a high degree of fluctuation over time. Figure 5.1 depicts the time series for total monthly U.S. tourist arrivals between January 1996 and November 2012. As can be seen from the graph, although the series increases over our sample period (with the exception of year 2001), its movements are dominated by seasonality. Since 2002 U.S. tourist arrivals exhibit a strong upward trend. The need to allocate resources for future growth is further evidence of the importance of developing accurate demand forecasting for investors, managers and policy makers in the tourism sector.
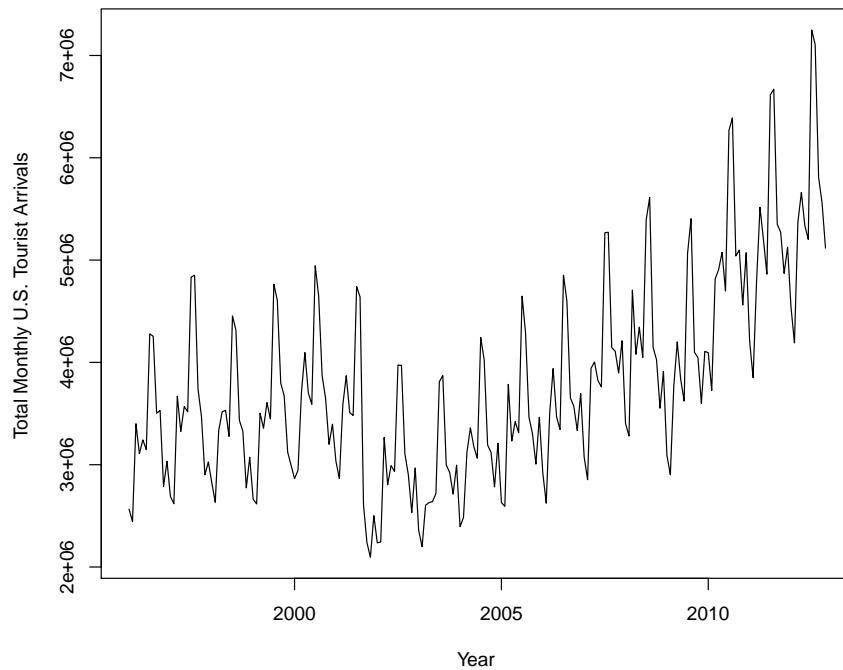
Fig. 5.1 Total monthly U.S. tourist arrivals time series (Jan. 1996 - Nov. 2012).

There are a number of components which define a good demand forecasting model for tourism management. Firstly, the forecasting model has to be able to pick up strong variations in tourist arrivals as most tourist demand time series show increasing fluctuations with seasons. Secondly, given the seasonal fluctuations, the measure of forecasting accuracy based on the forecasting error alone is not sufficient. It is important that the forecasting model is equally able to predict the actual direction of change. If not, investment decisions and the resources allocated to tourism could find themselves catering for a peak in demand but actually experiencing a trough. Thirdly, a tourism demand forecasting model needs to be efficient both in the short and long run. This is because long term investments are needed to be able to supply to the short term demand fluctuations. In this application all these aspects are considered as the SSA technique

is introduced for modelling and forecasting U.S. tourist arrivals, and compared with other forecasting models currently used to forecast tourism demand.

5.4.1 Literature Review

The existing literature on the forecasting of tourism demand is wide ranging both in terms of the different techniques employed and in terms of the different countries covered. A common theme in almost all of the papers also helps to explain the reasons behind this extensive interest in forecasting tourism demand. A large number of authors including Chan and Lim (2011), Chu (2008), Coshall and Charlesworth (2011) and, Goh and Law (2002) emphasise the importance of forecasting for investment and development planning in tourism. This message is re-enforced by authors, such as Gounopoulos et al. (2012) and Hui and Yuen (2002), who add that such forecasts are also important as a consequence of the vulnerability of tourism to large fluctuations in demand. Some authors also emphasise the importance of tourism to a particular economy to re-enforce the importance of accurately forecasting tourism demand. Examples include work by Jackman and Greenidge (2010) for Barbados, and Chu (2011) for Macau. Those readers seeking a detailed review of the literature, the paper by Song and Li (2008) covers 121 studies produced from 2000 to the date of publication. This review article offers a further reason for the sustained and extensive interest in forecasting tourism demand. They found that no single forecasting model outperforms all other in all possible situations. This implies that the literature is not only of importance but also in need of further research. A more recent review of forecasting and the closely related issue of tourism demand modelling is included in the paper by Song et al. (2012).

Perhaps the most common form of study is one that assesses the performance of one or more forecast techniques relative to a set of alternatives. Álvarez-Díaz

and Rosselló-Nadal (2010) examine forecasts of UK tourist arrivals in the Balerics, using meteorological variables. They compare the performance of an ARIMA model and a non-causal autoregressive neural network, finding that the latter performs better. Assaf et al. (2011) examine persistence and seasonality in data for tourist arrivals into Australia. They compare the performance of three different forecasting models, two standard methods using stationarity of degrees 0 and 1 and a model with fractional degrees of integration. Athanasopoulos and de Silva (2012), in a study of tourist arrivals in Australia and New Zealand propose a model which captures time varying seasonality within a vector innovation time series model. They produce evidence that this model offers greater forecast accuracy than a number of alternatives. Cho (2003) investigates three different techniques (exponential smoothing, univariate ARIMA and artificial neural networks) to forecast tourist arrivals in Hong Kong, finding the artificial neural networks forecasts to be the most accurate.

Chu (2008) explores fractionally integrated ARMA models in forecasting tourism arrivals in Singapore, observing that they perform well in comparison to more traditional ARIMA models. Chu (2011) compares a piecewise linear model with autoregressive trend, seasonal ARIMA and fractionally integrated ARMA models in forecasting tourism demand for Macau, concluding the piecewise linear model to be the most accurate. Likewise, Gil-Alana (2005) considers forecasts using monthly data for tourist arrivals into the US using a procedure combining unit and fractional integration in seasonal variation. He finds evidence of long memory and mean reverting behaviour. Goh and Law (2002) use data for Hong Kong tourist arrivals to compare forecasts from a stochastic nonstationary seasonality model (SARIMA) and an intervention component model (MARIMA) with a selection of eight other time series models. Their results suggest the SARIMA and MARIMA models to have the highest forecast accuracy of the models analysed.

Greenidge (2001) uses a structural time series model to provide and evaluate forecasts for tourism arrivals in Barbados. Jackman and Greenidge (2010) further explore the structural time series model for tourist arrivals in Barbados, finding that it produces more accurate forecasts than a number of alternatives. Hadavandi et al. (2011) present forecasts for tourism arrivals in Taiwan using a hybrid artificial intelligence model, involving a fuzzy rule-based system, which they found to be more accurate than a selection of three alternative approaches. Kim et al. (2011) consider the performance of prediction intervals for tourism arrivals into Hong Kong and Australia for a selection of time series forecasting models. They find an autoregressive bias corrected bootstrap model to perform best of those tested. Lim and McAleer (2001) analysed the performance of various different exponential smoothing models in forecasting tourist arrivals in Australia, concluding that using models expressed in first differences increased forecast accuracy. Shareef and McAleer (2007) evaluate the abilities of ARMA models to capture the effects of volatility in the time series of tourism arrivals in the Maldives. Song et al. (2010) focus on a different aspect of forecasting tourism demand - what is the appropriate measure of demand? Using data for Hong Kong they find use of tourism arrivals to be more affected by income in the country of origin and tourism expenditure to be more sensitive to prices. Wan et al. (2013), also using tourist arrival data for Hong Kong, assess the properties of disaggregated forecasts using a seasonal ARIMA model relative to aggregate forecasts. They find the sum of disaggregated forecasts to provide greater accuracy than an aggregate forecast.

A very closely related strand in the literature seeks to combine two or more forecasting models into a new hybrid model and to test whether this results in greater forecast accuracy. Andrawis et al. (2011) finds that, in forecasts of tourism arrivals into Egypt, combining short and long term forecasts improves accuracy compared to the individual forecasts. Cang (2011) examines tourism

arrivals into the U.K. and examines three different forecasting models - support vector neural networks, seasonal ARIMA and an exponential smoothing model. He finds that non-linear combinations of these models offer greater forecast accuracy than the individual specifications. Coshall and Charlesworth (2011) consider a number of forecasting models, both individually and in combination. Using data on UK outbound tourism they also find that forecast accuracy is improved by using a combination of forecasts.

Shen et al. (2008) focus on outward leisure tourism from the U.S. and examine seven different types of individual forecasting techniques. Their results also suggest that forecast accuracy is improved by combining forecasts. Shen et al. (2011) conduct a similar analysis of UK outward tourism, using seven different individual forecasting methods and six combinations. Again their findings suggest that forecast accuracy is improved by using combinations of forecasts. Song et al. (2011) develop a model to forecast Hong Kong tourist arrivals which combines a structural time series model with a time varying parameter one. They find that, relative to a number of time series models, their hybrid model exhibits greater forecast accuracy. Song et al. (2013), again with respect to tourism arrivals in Hong Kong, consider a model which combines quantitative forecasts which judgemental forecasting from an online survey. They find that adding a judgemental component improves forecast accuracy.

A number of papers consider the implications of shocks to one or more forecasting models of tourism demand. Gounopoulos et al. (2012) consider the forecasting of the impact on tourism arrivals in Greece of macro-economic shocks. They compare a number of different forecasting methods, finding an ARIMA model to be the most accurate and also develop a VAR model. Smeral (2010) examines the effects on forecasts of outbound travel of global recession for a sample of countries. Mao et al. (2010) use a cusp catastrophe model to forecast the rates of recovery of tourist arrivals in Taiwan from the SARS epidemic.

Their results suggest that tourism from China and the U.S. recovered quickly but that from Japan did not. In a similar vein Page et al. (2012) estimate the negative effect of the Swine flu epidemic on U.K. tourist arrivals using a time varying parameter model. Fourie and Santana-Gallego (2011) use a gravity model to estimate and predict the impact of mega-sports tourism events on tourist arrivals.

Studies which examine the determinants of demand for tourism are not analysis of forecasting models but are so closely related to the forecasting of tourism demand that they merit consideration. Chan and Lim (2011) analyse seasonality in New Zealand tourism demand using spectral analysis. They find different categories of inbound tourism share common cyclical behaviour. Naudé and Saayman (2005) consider the determinants of tourist arrivals in 43 African countries, finding tourism infrastructure and health risks to be of particular importance. Nelson et al. (2011) estimate a demand model for visitors to Hawaii from mainland U.S. Their results suggest home state income, airfares and (log) distance to be important. Seetanah et al. (2010) estimate tourism demand for South Africa using a gravity model. Their analysis suggests prices, level of development and common borders to all be important determinants. Seetaram (2010) uses dynamic panel cointegration to estimate demand elasticities for tourism arrivals into Australia, finding demand to be inelastic in the short run but elastic in the long run.

Volatility models are built upon an ARIMA model to which they add a second equation to explain the conditional variance. Coshall (2009) provides a good overview of these techniques and their application to forecasting tourism demand. The most commonly used specification is the GARCH model, developed by Bollerslev (1986). This adds to the ARIMA model an equation to explain the conditional variance. This equation models the current period conditional variance in terms of lagged squared residuals (capturing the short run impact of past shocks) and long term effects from lagged values of the condi-

tional variance. Extensions of the GARCH model include the TGARCH (which use dummy variables to model asymmetric shocks) and EGARCH models. For example, Kim and Wong (2006) use both the EGARCH and TGARCH models to provide forecasts of tourism demand in Korea with asymmetric responses to "news" shocks. Coshall (2009), in an application to UK outbound tourism, shows that forecasts using the EGARCH model can be combined with those from an exponential smoothing model such that the combined forecast is more accurate than either of the individual methods.

The use of SSA in the tourism sector was firstly evaluated by Beneki et al. (2012) via an application into signal extraction and forecasting of U.K. Tourist income. However, this application introduces SSA as a new model for forecasting tourism demand in the future. As noted above, there exist various different techniques which have been applied for forecasting tourism demand in the past. The performance of SSA forecasts are compared with the forecasting results from ARIMA (Automatic-ARIMA), Exponential Smoothing (ETS) and Neural Networks (nnetar). The ETS methodology gained popularity through its performance at the M3-competition. The state space framework which underlies the new developments in ETS is widely applicable, and like ARIMA, underpins forecasts with a sound stochastic model (Hyndman et al., 2002). Neural networks have frequently been adopted in tourism demand forecasting as previously mentioned. A further key feature is that used in this application is the most basic version of SSA-V with optimal choices. Given the choice of forecasting methods, the forecasting accuracy of both parametric and nonparametric time series analysis and forecasting techniques are compared. Unlike parametric forecasting techniques, nonparametric techniques are not bound by any of the usual assumptions such as stationarity and normality. As such, nonparametric models are better able to provide a true approximation of the real situation. However, it is important to note that this application does not intend on showing

that SSA is the universally best model for forecasting tourist arrivals. Instead, the aim is the introduction of SSA as an alternative method, and further research is required to compare SSA's performance against many other forecasting techniques.

5.4.2 The Data

Used for application purposes is the monthly U.S. tourist arrivals data from January 1996 to November 2012 (203 observations) obtained via the U.S. Department of Commerce: Office of Travel and Tourism Industries[1]. Table 5.1 provides some descriptives for the data. According to the data, average total monthly tourist arrivals into the U.S. between January 1996-November 2012 were 3,798,000. The maximum number of tourist arrivals during the sample period of concern was recorded at 7,249,000 in July 2012 and the minimum 2,096,000 (in November 2001). By region the lowest average monthly tourist arrivals into U.S. were recorded from Africa whilst Canada accounts for an average of 1,346,000 tourist arrivals, the highest influx of tourists into U.S. from a single country. The skewness statistic indicates that all time series analysed in this study are in fact skewed and not normally distributed. An analysis of the kurtosis suggests that all the series have Platykurtic distributions except for Italy which has a Leptokurtic distribution. Accordingly, this information tells us that the Italian time series for tourist arrivals into U.S. has a high probability for extreme values with thicker tails and values concentrated around the mean whilst all other time series for U.S. tourist arrivals have a lesser probability for extreme values in comparison to a normal distribution and consist of values which have a wider spread around the mean. In order to confirm the information provided through the skewness and kurtosis statistics, the data was tested for normality

---

[1]http://travel.trade.gov/research/monthly/arrivals/

using the Shapiro-Wilk test. Accordingly, it was found that Western Europe, Total Overseas, Asia and Central America were in fact normally distributed at a *p*-value of 0.05. The last column in Table 5.1 shows the seasonal R-square. This is obtained as the conventional R-square in a regression of the first difference series against twelve monthly dummy variables. The R-square for Canada is the largest and accounts for 92% of total variation in the series. It is noteworthy that with the single exception of Hong Kong, monthly dummy variables account for over 60% of the variation for each country. Thus it is possible to conclude that seasonality is generally strong for these series.

Table 5.1 Descriptive statistics: U.S. tourist arrivals (Jan. 1996 - Nov. 2012).

Series	Mean	Min.	Max.	Std. Dev.	Skew.	Kurtosis	Seasonal R^2
Total Arrivals	3798000	2096000	7249000	994944	0.86	0.63	0.88
Arrivals by country							
Canada	1346000	727300	2945000	417184	1.07	1.24	0.92
Mexico	491100	67960	1668000	338299	1.25	0.74	0.67
Total Overseas	1961000	1119000	3089000	382831	0.34	-0.06	0.77
Western Europe	859700	418800	1320000	187797	0.14	-0.50	0.78
Eastern Europe	39000	17610	76360	11875	0.69	-0.15	0.78
Asia	550500	246500	934300	106414	0.22	0.71	0.71
Middle East	51600	22930	120200	17996	1.09	1.21	0.63
Africa	22870	7869	48080	6863	0.63	0.51	0.87
Oceania	65190	28090	165600	23470	1.26	1.81	0.87
South America	215200	98580	420300	68877	0.62	0.06	0.67
Central America	59510	29730	91860	12097	0.29	-0.14	0.83
Caribbean	97440	48330	191100	31712	1.05	0.51	0.89
France	86290	36920	201800	31954	1.20	1.39	0.83
Germany	136800	54920	235600	39695	0.24	-0.80	0.71
Italy	51460	17170	157400	23127	1.88	4.65	0.85
Netherlands	41180	20340	90430	12554	1.26	2.23	0.83
Spain	36260	13110	104600	16651	1.40	2.15	0.67
Sweden	25820	11070	51560	7680	0.84	0.96	0.75
Switzerland	29090	13270	74220	10514	1.24	2.22	0.82
United Kingdom	338400	164300	475400	64735	-0.46	-0.34	0.74
Japan	331200	141600	549100	80225	0.36	-0.21	0.66
South Korea	62490	19510	130300	22956	0.78	0.24	0.78
PRC & Hongkong	46480	11480	207000	28966	2.63	8.62	0.43
ROC (Taiwan)	27830	9451	63400	10223	0.90	0.70	0.70
Australia	51380	21000	142400	20462	1.41	2.17	0.72
Argentina	30780	9279	64240	13845	0.25	-1.00	0.60
Brazil	65960	18680	171000	34633	1.11	0.69	0.72
Colombia	31810	11110	74670	12050	0.79	0.18	0.83
Venezuela	39330	15780	86160	14841	0.93	0.51	0.83

Next the U.S. tourist arrivals series are tested for unit root problems as certain external shocks such as recessions (for example) are infamous for creating

structural breaks and making a time series nonstationary in mean and variance. Table 5.2 reports the findings from the Bai and Perron (2003) test for structural breaks in the U.S. tourist arrivals series. Whilst analysing the causes and reasons behind these structural breaks are beyond the mandate of this application, certain interesting observations are outlined. Firstly, it is evident that based on the Bai and Perron (2003) test, the time series relating to tourist arrivals from the Caribbean is the only series that has not been affected by structural breaks. Secondly, except for Canada, Mexico, Africa, Central America, Germany, Italy, Spain, United Kingdom, South Korea, PRC & Hongkong, Australia and Colombia, all other time series considered in this study are affected by a structural break in the year 2001. Furthermore, the Bai and Perron (2003) test indicates there has been a delayed impact of the 2008 recession on U.S. tourist arrivals with all series reporting structural breaks in 2010 with the exception of Mexico, Western Europe, Africa, Germany, Italy, Netherlands, Spain, United Kingdom, Japan, ROC (Taiwan), Colombia and Venezuela. Finally, in terms of U.S. tourist arrivals by country of origin, the most number of structural breaks visible in a time series is seen in tourist arrivals from Brazil.

Table 5.3 presents the model parameters (SSA choices) for each of the forecasting techniques considered in this study for forecasting total U.S. tourist arrivals at horizons of $h = 1, 3, 6, 12, 24$ and 36 months ahead. It is important to note that each of the techniques have chosen the model parameters (SSA choices) automatically using the respective algorithms (as explained in Chapters 2.4 and 4.3.1) to provide the best possible modelling and forecast for U.S. tourist arrivals.

Considered next are the SSA-V decompositions which is an integral part of the SSA process. The weighted correlation (w-correlation) statistic is used to show the appropriateness of the various decompositions achieved by SSA (see, Table 4.3 and Table 4.9). As mentioned in Golyandina et al. (2001), the w-

Table 5.2 Break points in U.S. tourist arrivals time series.

Series	Structural Break
Total Arrivals	2001(9), 2004(3), 2007(2), 2010(2)
Arrivals by country	
Canada	2006(2), 2010(2)
Mexico	1998(6), 2006(3), 2009(12)
Total Overseas	2001(8), 2004(3), 2007(4), 2010(4)
Western Europe	2001(8), 2004(2), 2007(6)
Eastern Europe	2001(8), 2004(5), 2007(4), 2010(5)
Asia	2001(8), 2004(4), 2010(4)
Middle East	1999(2), 2001(8), 2006(5), 2010(5)
Africa	2008(4)
Oceania	2001(9), 2004(3), 2007(3), 2010(4)
South America	2001(8), 2007(5), 2010(5)
Central America	1998(6), 2001(8), 2007(4)
Caribbean	No structural break in series.
France	2001(8), 2007(3), 2010(3)
Germany	2000(10), 2007(2)
Italy	2007(6)
Netherlands	2001(8), 2007(3)
Spain	2007(5)
Sweden	2001(6), 2004(2), 2007(2), 2010(5)
Switzerland	2001(7), 2007(3), 2010(5)
United Kingdom	1998(6), 2008(12)
Japan	2001(8)
South Korea	2005(4), 2010(4)
PRC & Hongkong	2007(4), 2010(5)
ROC (Taiwan)	2001(8)
Australia	2005(4), 2010(4)
Argentina	2001(8), 2006(12), 2010(4)
Brazil	1999(1), 2001(7), 2005(4), 2007(11), 2010(5)
Colombia	2009(5)
Venezuela	2001(12), 2007(6)

Table 5.3 Forecasting model parameters for total U.S. tourist arrivals.

h	ARIMA	ETS(α, γ, σ)	NN(p, P, k)	SSA(L, r)
1	order(2,0,1)seasonal(1,1,2)	$(0.55, 0.18, 0.05)^M$	NNAR(2,1,1)	(38,17)
3	order(2,0,1)seasonal(1,1,2)	$(0.55, 0.18, 0.05)^M$	NNAR(2,1,1)	(25,14)
6	order(2,0,1)seasonal(1,1,2)	$(0.55, 0.18, 0.05)^M$	NNAR(2,1,1)	(29,21)
12	order(2,0,1)seasonal(1,1,2)	$(0.55, 0.18, 0.05)^M$	NNAR(2,1,1)	(15,6)
24	order(2,0,1)seasonal(1,1,2)	$(0.55, 0.18, 0.05)^M$	NNAR(2,1,1)	(40,25)
36	order(2,0,1)seasonal(1,1,2)	$(0.55, 0.18, 0.05)^M$	NNAR(2,1,1)	(48,6)

Note:^{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. L is the window length and r is the number of eigenvalues.

correlation statistic which shows the dependence between two time series 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, $\parallel Y_N^{(i)} \parallel_w = \sqrt{\left(Y_N^{(i)}, Y_N^{(i)}\right)_w}$, $\left(Y_N^{(i)}, Y_N^{(j)}\right)_w =$

$\sum_{k=1}^{N} w_k y_k^{(i)} y_k^{(j)}$ $(i, j = 1, 2)$, $w_k = \min\{k, L, N-k\}$ (here, assume $L \leq N/2$).

Accordingly, if the *w*-correlation between two reconstructed components are close to 0, this implies that the corresponding series are *w*-orthogonal and in turn we know the two components are well separable (Hassani et al., 2009). Table 5.4 presents the *w*-correlations for all the decompositions by comparing the two components of signal and noise. Here, used as signal is the reconstructed series containing optimal *r* components whilst the remaining *r* (which does not belong to the reconstruction) are selected as noise. The results indicate that all *w*-correlations are close to 0 which in turn suggests that a sound decomposition has been achieved using the automated & optimized SSA-V forecasting algorithm explained in Section 4.3.1. In other words, these *w*-correlations indicate that the newly proposed SSA-V forecasting algorithm works exceedingly well at separating the noise from the signal.

Table 5.4 *W*-correlations between signal and residuals for U.S. tourist arrivals.

Series	1	3	6	12	24	36
Total U.S. tourist Arrivals	0.007	0.009	0.009	0.012	0.008	0.009
U.S. tourist Arrivals by country						
Canada	0.013	0.010	0.010	0.028	0.010	0.012
Mexico	0.020	0.020	0.021	0.047	0.032	0.035
Total Overseas	0.009	0.009	0.014	0.014	0.008	0.006
Western Europe	0.010	0.014	0.015	0.019	0.024	0.012
Eastern Europe	0.020	0.016	0.014	0.015	0.022	0.020
Asia	0.008	0.008	0.008	0.017	0.007	0.006
Middle East	0.027	0.047	0.044	0.029	0.022	0.024
Africa	0.019	0.020	0.015	0.031	0.013	0.010
Oceania	0.010	0.009	0.014	0.018	0.007	0.007
South America	0.012	0.019	0.023	0.016	0.020	0.023
Central America	0.013	0.016	0.014	0.021	0.012	0.016
Caribbean	0.021	0.021	0.031	0.051	0.034	0.019
France	0.014	0.027	0.040	0.015	0.014	0.015
Germany	0.015	0.015	0.014	0.017	0.017	0.017
Italy	0.026	0.026	0.026	0.016	0.035	0.024
Netherlands	0.016	0.018	0.018	0.018	0.027	0.014
Spain	0.030	0.014	0.031	0.027	0.016	0.027
Sweden	0.012	0.012	0.012	0.011	0.012	0.017
Switzerland	0.024	0.016	0.021	0.017	0.020	0.020
United Kingdom	0.013	0.016	0.015	0.013	0.012	0.016
Japan	0.009	0.015	0.008	0.009	0.07	0.012
South Korea	0.016	0.016	0.012	0.016	0.016	0.012
PRC & Hongkong	0.025	0.051	0.022	0.030	0.025	0.022
ROC (Taiwan)	0.019	0.031	0.025	0.025	0.015	0.015
Australia	0.011	0.011	0.011	0.011	0.011	0.011
Argentina	0.028	0.010	0.046	0.029	0.007	0.010
Brazil	0.025	0.023	0.026	0.027	0.030	0.027
Colombia	0.022	0.023	0.012	0.012	0.019	0.038
Venezuela	0.026	0.025	0.026	0.046	0.022	0.021

The U.S. tourist arrivals series exhibits several seasonal patterns. In order to illustrate SSA's capabilities at extracting various seasonal patterns in U.S. tourist arrivals, also presented via Figure 5.2 as an example, is the in-sample

decomposition of total U.S. tourist arrivals at $h = 1$ step ahead. Firstly, it is observable that the extracted trend in U.S. tourist arrivals which corresponds with the total arrivals pattern and clearly shows the general trend of increasing and decreasing tourist arrivals over time. Also interesting is the difference between the four month and twelve month seasonal components. The 4 month seasonal component is increasing over time whilst the 12 month seasonal component is seen to be decreasing over time. Furthermore, there is more fluctuation in the 4 month seasonal component of total U.S. tourist arrivals in comparison to the 12 month component.
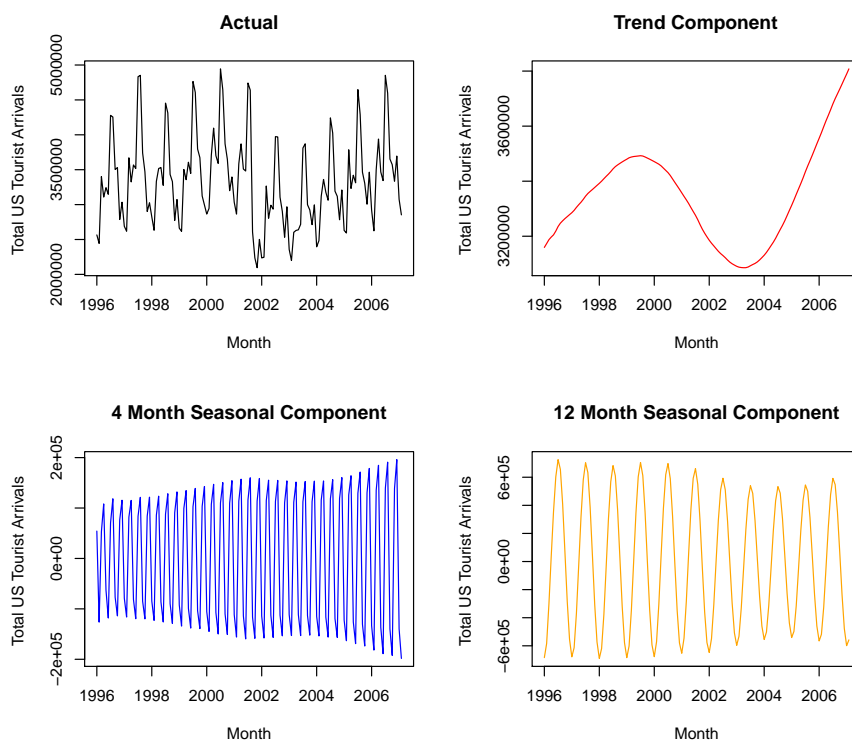

Fig. 5.2 In-sample SSA decomposition of total monthly U.S. tourist arrivals at $h = 1$ step ahead.

5.4.3   Empirical Results

The application considers $\frac{2}{3}^{rd}$ of the data as in-sample for model training and testing, and set aside $\frac{1}{3}^{rd}$ of the data as out-of-sample for evaluating the forecasting accuracy. The data was forecasted at horizons of $h = 1, 3, 6, 12, 24$ and 36 steps ahead which corresponds to 1, 3, 6, 12, 24 and 36 months ahead forecasts. These forecasting horizons have been considered because for the tourism industry, horizons beyond 12 months are considered to be long term. Moreover, both short and long run forecasts are vital for this sector as a country needs to be geared to accommodate the tourists and planning of large scale building works or the purchase of new aircrafts for example would require managerial decisions to be made well in advance. Therefore, this application is effectively evaluating the performance of the forecasting models both in the short and long run in terms of obtaining forecasts for U.S. tourist arrivals. Initially total U.S. tourist arrivals are analyzed. Table 5.5 reports the RMSE and MAPE results for the out-of-sample forecasts of total U.S. tourist arrivals using SSA, ARIMA, ETS and NN. In order to ensure the parametric models are correctly specified, a Ljung-Box test was carried out on the residuals for autocorrelation and the results indicated that the residuals are independently distributed at a p-value of 0.05, and are thus not autocorrelated.

Table 5.5 Out-of-sample RMSE(MAPE) results for total U.S. tourist arrivals.

h	ARIMA	ETS	NN	SSA	$\frac{SSA-V}{ARIMA}$	$\frac{SSA-V}{ETS}$	$\frac{SSA-V}{NN}$
1	601512 (9%)	760599 (13%)	1147080 (19%)	242601(4%)	0.40*	0.32*	0.21*
3	720751 (11%)	723556 (13%)	1124242 (19%)	316049(6%)	0.44*	0.44*	0.28*
6	738630 (12%)	1037666 (20%)	1180780 (19%)	445614(8%)	0.60*	0.43*	0.38*
12	937129 (14%)	1097366 (17%)	1385339 (23%)	517912(9%)	0.55*	0.47*	0.37*
24	1136616 (19%)	1300442 (20%)	1780513 (30%)	526323(9%)	0.46*	0.40*	0.30*
36	1002685 (17%)	1149585 (18%)	1684799 (24%)	605448(9%)	0.60*	0.53*	0.36*
Average	856221 (14%)	1011536 (17%)	1383792 (22%)	442325(8%)	0.52	0.44	0.32

Note:* indicates results are statistically significant based on Diebold-Mariano at $p = 0.05$.

Firstly, based on the MAPE criterion reported in Table 5.5, it is clear that the Neural Network model is the worst performer at all horizons with an overall

average MAPE of 22% at forecasting total U.S. tourist arrivals. Interestingly the SSA technique is the only model which reports MAPE values below 10% at all horizons and is in turn the model providing the most accurate forecasts for total U.S. tourist arrivals with a comparatively low average MAPE of 8%. Based on the MAPE one is also able to identify that the ARIMA model is the second best model for forecasting total U.S. tourist arrivals as its average MAPE of 14% is lower than the ETS model's average MAPE of 17%. Moreover, it is interesting to note that the SSA model's MAPE remains approximately constant over the forecasting horizons of $h = 12$, 24 and 36 months ahead, and thereby portrays SSA's capabilities of providing comparatively stable and more accurate forecasts in the long run. The remainder of the analysis focusses on the RMSE criterion for evaluating forecast accuracy.

It is evident from Table 5.5 that based on the RMSE criterion, SSA outperforms ARIMA, ETS and Neural Networks comfortably by recording the lowest forecasting error for total U.S. tourist arrivals at all horizons. The RRMSE statistic shows that SSA is 60%, 56%, 40%, 45%, 54% and 40% better than ARIMA at forecasting total U.S. tourist arrivals at $h = 1, 3, 6, 12, 24$ and 36 months ahead respectively. Likewise, in comparison to ETS, SSA is 68%, 56%, 57%, 53%, 60% and 47% better at $h = 1, 3, 6, 12, 24$ and 36 steps ahead respectively. Analyzed finally are the forecasting results between SSA and the Neural Network model. Accordingly it is possible to conclude that the SSA model is 79%, 72%, 62%, 63%, 70% and 64% better than the feed-forward Neural Network model at $h = 1, 3, 6, 12, 24$ and 36 months ahead respectively.

In order to ensure the results reported are not chance occurrences, they were further tested for statistical significance using the modified Diebold-Mariano test found in Harvey et al. (1997). The test results indicate that all the RRMSE results are statistically significant at all horizons and thus provides concrete evidence for the inferences made via the application. Finally, from Table 5.5 one

can infer that when forecasting total U.S. tourist arrivals, on average, the SSA model is 48% better than ARIMA, 56% better than ETS and 68% better than Neural Networks based on the forecasting accuracy. The results from Table 5.5 also show that on average, ARIMA provides a better forecasting accuracy in comparison to ETS and Neural Networks for U.S. tourist arrivals both in the short and long run, and is therefore chosen to be the second best model in general for this purpose. The feed-forward Neural Network model with one hidden layer provides the least favourable forecasts for total U.S. tourist arrivals.

The impact of the 9/11 terrorist attack on U.S. soil is clearly identifiable in Figure 5.1. The breakpoints test carried out earlier has confirmed this particular structural break occurred in September 2001. As economic literature provides evidence of such breaks impacting unit root tests such as KPSS, the ARIMA and SSA models are further tested for robustness to this break. Accordingly, the data were re-modelled by considering data post September 2001 to ascertain whether this major break in the series has a significant impact on ARIMA or SSA's modelling capabilities. The out-of-sample forecasting results from the re-modelling is presented in Table 5.6. As appears from these results both models now perform better than previously in the absence of this break. However, it is clear that SSA continues to dominate with the lowest RMSE and MAPE results at all horizons in comparison to ARIMA thus provides further evidence for the reliability of the results presented in Table 5.5.

Thereafter, the Direction of Change (DC) criterion was used to evaluate the extent to which the forecasts from all models are able to predict the actual direction of change in total U.S. tourist arrivals. Table 5.7 presents the DC results. Firstly, it is evident that only three outcomes are in fact statistically significant for DC. However, based on the criterion itself one could infer that SSA provides a more accurate prediction of direction of change in comparison to ARIMA at all horizons when forecasting total U.S. tourist arrivals, and on average, SSA is

Table 5.6 Out-of-sample RMSE(MAPE) results for total U.S. tourist arrivals (adjusted for the 9/11 breakpoint).

h	ARIMA	SSA-V	$\frac{SSA-V}{ARIMA}$
1	648372 (10%)	289642 (5%)	0.45*
3	683034 (11%)	354762 (6%)	0.52*
6	834528 (12%)	422322 (7%)	0.51*
12	846094 (11%)	345101 (5%)	0.41*
24	827373 (11%)	388009 (5%)	0.47*
36	813722 (10%)	447459 (6%)	0.55*
Average	775521 (11%)	374549 (6%)	0.48

Note:* indicates results are statistically significant based on Diebold-Mariano at $p = 0.05$.

able to provide a 83% accurate direction of change prediction whilst ARIMA can only provide a 63% accurate prediction of the direction of change. Likewise, in comparison to both ETS and Neural Networks, SSA provides a better prediction of the direction of change at all horizons. However ETS outperforms the ARIMA model in terms of DC at $h = 3$ and 24 months ahead and the DC predictions of the NN model is better than ETS at $h = 12$ and 24 steps ahead. Furthermore, at 36 steps ahead the SSA model obtains 100% accurate DC predictions whilst ARIMA is able to report a significant 91% accuracy. Thus, it is clear that the SSA model stands out as the most superior model for forecasting total U.S. tourist arrivals at all horizons based on the RMSE, RRMSE and DC criterions in comparison to ARIMA, ETS and Neural Network models. Furthermore, it is clear that the SSA model can pick up both short and long run fluctuations in total U.S. tourist arrivals comparatively better than ARIMA, ETS and Neural Networks.

As an example of the out-of-sample forecasting capabilities of the selected models, and also to show the accuracy of the DC results, presented in Figure 5.3 is a graphical representation of the forecasting results at $h = 24$ steps ahead for total U.S. tourist arrivals. The choice of this particular horizon as the example is for the following reasons. First and foremost, it is well known that forecasting any variable becomes increasingly difficult as the horizon increases. In this

Table 5.7 Direction of change results for total U.S. tourist arrivals forecasts.

h	ARIMA	ETS	NN	SSA-V
1	0.74*	0.57	0.48	0.87**
3	0.70*	0.73*	0.57	0.85**
6	0.67*	0.63*	0.56	0.81*
12	0.47	0.36	0.45	0.66*
24	0.30	0.52	0.63*	0.78*
36	0.91**	0.56	0.56	1.00**
Average	0.63	0.56	0.54	0.83

Note:** indicates results are statistically significant based on a t-test at $p = 0.05$.
* indicates results are statistically significant based on a t-test at $p = 0.10$.

case, $h = 24$ steps ahead represents a long run forecast horizon considered in this application and thus can show the reader how well or poorly the models fare at predicting total U.S. tourist arrivals in the long run. Secondly, this also happens to be the forecasting horizon at which SSA reports 78% accuracy in terms of the correct direction of change prediction whilst ARIMA reports 30% accuracy which is also the lowest ARIMA recorded over all horizons considered in this application. As such, this plot can show the reader a further clear difference between the best (SSA) and second best (ARIMA) forecasts for total U.S. tourist arrivals. Thirdly, this is the only horizon at which NN reports a statistically significant direction of change prediction at 63%. As such this plot also enables to notify the reader of the fact that the direction of change criterion should be taken into consideration alongside a loss function when determining which model is best for forecasting. In fact if one was to pick the second best model based on the DC criterion alone, then in this case they would opt for NN whilst in terms of the loss function it is the worst performing model reporting the highest MAPE of 30% across all horizons considered in this application. Whilst NN would provide a better DC prediction than ARIMA in this case, relying on NN forecasts for planning and decision making would result in major unproductive resource allocations given that the actual forecasts themselves report very high

errors. This particular figure also clearly indicates the very poor nature of forecasts achieved via the NN model. In terms of Figure 5.3 itself, it is evident that both ETS and NN models experience great difficulty in picking up the seasonal fluctuations seen in the U.S. tourist arrivals time series and that the NN model is indeed the worst performer in this case. The results from both Tables 5.5 and 5.6 proves that as the horizon increases from 1 month ahead to 24 months ahead, the forecasting performance of the parametric model (ARIMA), ETS and NN worsens immensely in comparison to that of the nonparametric model of SSA.

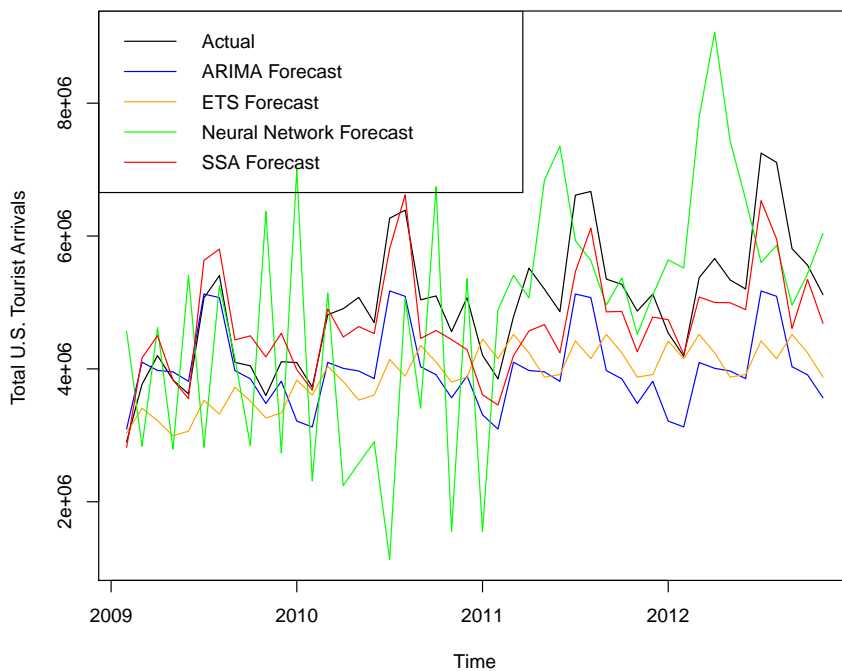
Fig. 5.3 $h = 24$ months ahead forecast for U.S. tourist arrivals (Feb. 2009 - Nov. 2012).

The positive outcome when forecasting Total U.S. tourist arrivals using the new SSA algorithm inspired considering same for forecasting U.S. tourist arrivals by country of origin. The total U.S. tourist arrivals forecasting results show ARIMA to be the second best forecasting model in comparison to SSA,

ETS and Neural Networks. As such, here ARIMA is employed as a benchmark as it is evident that ETS and feed-forward Neural Networks cannot provide accurate forecasts in this case. Presented in Tables 4.8 and 5.9 are the ARIMA parameters and SSA-V choices which were used for forecasting U.S. tourist arrivals by country of origin. Once again the correct specification of the models were ensured via a Ljung-Box test for the independent distribution of residuals. Where the residuals were not found to be independently distributed the model parameters we redefined to ensure the model specification was valid. In most cases the test results indicated that the residuals were white noise at a p-value of 0.05, and that no further model review was required.

Table 5.10 reports the results for out-of-sample forecasting of U.S. tourist arrivals by country of origin. The RRMSE criterion shows that, SSA outperforms ARIMA at forecasting U.S. Tourist arrivals at all horizons for all countries of origin with the exception of Mexico at $h = 3$ steps ahead. Furthermore, it is clear from the results that on average, SSA is 53%, 49%, 44%, 47%, 46% and 41% better than ARIMA at horizons of $h = 1, 3, 6, 12, 24$ and 36 months ahead respectively for forecasting U.S. tourist arrivals by individual country of origin.

These results prove that by employing SSA to analyse and forecast the monthly U.S. tourist arrivals data by country of origin we can obtain significantly more accurate forecasts than those possible with ARIMA for both short and long term fluctuations in tourist arrivals into the U.S. from each country. The results are tested further for statistical significance. The testing showed that except for tourist arrivals from Mexico, every other forecasting result obtained in this study is statistically significant. This suggests that when forecasting tourist arrivals from Mexico there is no difference between the forecasting accuracy of the ARIMA and SSA models.

Interestingly, when forecasting U.S. tourist arrivals from Mexico, the optimal SSA choice for the number of eigenvalues, r is $r = 1$ at horizons of 1, 3 and 6

Table 5.8 ARIMA model parameters for U.S. tourist arrivals by country of origin.

Series	ARIMA	Series	ARIMA
Canada	(0,0,1)(0,1,2)	Mexico	(1,1,3)(0,1,2)
Total Overseas	(0,1,1)(2,0,2)	Western Europe	(1,0,0)(2,0,1)
Eastern Europe	(2,0,1)(1,1,2)	Asia	(0,1,0)(2,0,1)
Middle East	(2,0,1)(2,0,2)	Africa	(2,0,3)(2,1,2)
Oceania	(3,0,3)(1,1,1)	South America	(0,1,2)(2,0,1)
Central America	(2,1,1)(2,1,1)	Caribbean	(0,0,1)(1,1,2)
France	$(1,1,1)(2,0,2)^{\dagger}$	Germany	(2,1,3)(2,0,2)
Italy	$(2,0,2)(1,1,2)^{\dagger}$	Netherlands	(4,0,4)(2,1,2)
Spain	(3,0,3)(1,1,1)	Sweden	(2,1,1)(1,1,2)
Switzerland	(5,1,4)(2,0,2)	UK	(2,1,4)(2,0,1)
Japan	(2,1,2)(2,0,1)	South Korea	$(1,0,1)(2,1,0)^{\dagger}$
PRC & Hongkong	(1,0,0)(2,0,2)*	ROC	(4,1,2)(2,0,2)
Australia	(4,1,5)(0,1,1)	Argentina	(1,1,1)(2,0,1)
Brazil	(1,1,2)(2,0,1)	Colombia	(2,0,4)(2,1,0)
Venezuela	(3,1,1)(1,1,1)		

Note:[†] ARIMA with drift. * ARIMA with non-zero mean.

steps ahead. This in turn means that the SSA model is relying on the trend alone to forecast future data points for Mexico. As such it is important to briefly comment on this fact. For this purpose, shown in Figure 5.4 are the time series for Mexico and three other time series which were found to have structural breaks as per Table 5.2. Upon closer analysis it is clear that whilst all four time series shown here are affected by structural breaks, the time series for Mexico shows signs of a major structural break shifting U.S. tourist arrivals from Mexico starting December 2009. The magnitude of this break has implications on SSA's modelling capabilities especially as this particular SSA-V forecasting algorithm does not incorporate change point detection methods. Moreover, it is clear from Figure 5.4 that U.S. tourist arrivals from Mexico differs from the other nations in terms of seasonality, as Mexico does not illustrate a strong seasonality. It is pertinent to keep in mind that in line with ensuring equality between the other forecasting models adopted in this application, used here is the most basic version of SSA with optimal choices for the purpose of forecasting U.S. tourist

Table 5.9 U.S. tourist arrivals by country of origin - SSA-V choices (L, r).

Series	1	3	6	12	24	36
Canada	(22,16)	(28,19)	(28,19)	(16,9)	(33,20)	(36,15)
Mexico	(5,1)	(5,1)	(4,1)	(51,3)	(39,9)	(49,26)
Total Overseas	(28,16)	(29,16)	(20,10)	(18,11)	(38,25)	(50,29)
Western Europe	(29,17)	(23,14)	(23,12)	(19,14)	(21,13)	(28,21)
Eastern Europe	(14,13)	(22,11)	(23,11)	(20,14)	(18,13)	(49,5)
Asia	(29,23)	(25,22)	(29,23)	(23,11)	(31,28)	(49,40)
Middle East	(24,15)	(15,13)	(17,13)	(22,18)	(44,36)	(38,15)
Africa	(18,14)	(17,14)	(24,20)	(14,12)	(47,17)	(24,16)
Oceania	(39,25)	(42,27)	(31,19)	(34,12)	(33,27)	(33,27)
South America	(27,14)	(23,15)	(16,12)	(26,15)	(46,28)	(35,24)
Central America	(29,17)	(29,19)	(26,17)	(29,25)	(47,24)	(46,27)
Caribbean	(24,20)	(24,20)	(24,11)	(18,12)	(24,12)	(46,26)
France	(15,13)	(30,14)	(25,12)	(43,31)	(23,20)	(40,21)
Germany	(32,10)	(25,8)	(25,9)	(32,15)	(24,12)	(24,12)
Italy	(44,15)	(34,15)	(34,15)	(57,27)	(18,14)	(30,23)
Netherlands	(36,14)	(37,14)	(26,19)	(26,19)	(22,10)	(32,11)
Spain	(28,8)	(12,6)	(24,8)	(14,9)	(14,3)	(14,9)
Sweden	(39,11)	(39,11)	(39,11)	(38,15)	(23,20)	(24,15)
Switzerland	(15,12)	(44,38)	(16,13)	(31,21)	(26,17)	(26,17)
UK	(24,14)	(22,14)	(32,24)	(51,38)	(41,34)	(47,14)
Japan	(31,25)	(28,9)	(47,19)	(23,21)	(47,34)	(39,9)
South Korea	(32,18)	(27,17)	(28,25)	(31,21)	(39,15)	(50,36)
PRC and Hongkong	(40,18)	(16,13)	(41,21)	(25,15)	(50,21)	(42,34)
ROC	(40,21)	(40,31)	(37,33)	(37,33)	(37,16)	(37,16)
Australia	(55,19)	(37,21)	(37,12)	(36,12)	(49,33)	(48,33)
Argentina	(23,15)	(30,26)	(15,13)	(17,15)	(41,39)	(41,40)
Brazil	(26,15)	(14,11)	(46,12)	(39,24)	(39,22)	(50,12)
Colombia	(29,15)	(29,16)	(39,11)	(36,11)	(27,23)	(19,10)
Venezuela	(30,15)	(28,15)	(26,15)	(18,15)	(48,15)	(37,17)

arrivals.

5.4.4 Discussion

The starting point of this application, as with many other authors, was the importance of accurate forecasts of tourism demand to investors, managers and policy makers. The existence of a high degree of seasonality in tourism demand not only increases this need, but also creates a need for forecasting techniques

Table 5.10 Forecasting results for U.S. tourist arrivals by country of origin.

Origin	$\frac{SSA-V}{ARIMA}$					
	1	3	6	12	24	36
Canada	0.27*	0.32*	0.40*	0.37*	0.30*	0.36*
Mexico	0.98	0.96	1.07	0.99	0.93	0.77**
Total Overseas	0.44*	0.48*	0.42*	0.48*	0.33*	0.43*
Western Europe	0.46*	0.50*	0.53*	0.47*	0.43*	0.44*
Eastern Europe	0.34*	0.37*	0.42*	0.41*	0.42*	0.39*
Asia	0.54*	0.68*	0.72	0.91	0.80*	0.91
Middle East	0.55*	0.42*	0.47*	0.37*	0.38*	0.46*
Africa	0.28*	0.39*	0.45*	0.36*	0.26*	0.24*
Oceania	0.40*	0.43*	0.51*	0.53*	0.60*	0.75**
South America	0.43*	0.45*	0.56*	0.50*	0.49*	0.82
Central America	0.44*	0.45*	0.52*	0.46*	0.34*	0.46*
Caribbean	0.34*	0.38*	0.43*	0.34*	0.49*	0.61*
France	0.36*	0.45*	0.42*	0.36*	0.52*	0.42*
Germany	0.60*	0.51*	0.64*	0.61*	0.64*	0.60*
Italy	0.31*	0.37*	0.41*	0.38*	0.35*	0.44*
Netherlands	0.48*	0.53*	0.47*	0.44*	0.44*	0.43*
Spain	0.60*	0.78*	0.76**	0.62*	0.65*	0.93
Sweden	0.53*	0.62*	0.72*	0.69*	0.62*	0.47*
Switzerland	0.48*	0.50*	0.54*	0.50*	0.48*	0.42*
United Kingdom	0.52*	0.49*	0.61*	0.65*	0.72**	0.92
Japan	0.62**	0.83*	0.82	0.71**	0.66*	0.96
South Korea	0.48*	0.49*	0.73*	0.79*	0.88	0.91
PRC & Hongkong	0.51*	0.52*	0.56*	0.47*	0.73**	0.64*
ROC (Taiwan)	0.50*	0.44*	0.48*	0.58*	0.50*	0.40*
Australia	0.44*	0.45*	0.48*	0.49*	0.61*	0.59*
Argentina	0.54*	0.62*	0.75*	0.64*	0.61*	0.59*
Brazil	0.53*	0.53*	0.58*	0.53*	0.49*	0.49*
Colombia	0.34*	0.38*	0.41*	0.41*	0.35*	0.66*
Venezuela	0.42*	0.34*	0.34*	0.44*	0.53*	0.52*
Average	0.47	0.51	0.56	0.53	0.54	0.59

Note:* indicates results are statistically significant based on Diebold-Mariano at $p = 0.05$.
** indicates statistical significance at $p = 0.10$.

that cope well with this seasonality in time series. Singular Spectrum Analysis is introduced as a new model for forecasting inbound U.S. tourist arrivals. The U.S. tourist arrivals time series' are analyzed in total and by country of origin.

This analysis compared the forecasting accuracy of the newly proposed technique, the automated and optimized SSA-V model, with the forecasting accuracy of the several different widely used forecasting models. These include an optimized version of ARIMA, known as Automatic-ARIMA, an Exponential Smoothing model known as ETS and a feed-forward Neural Network model known as nnetar. Automatic-ARIMA, ETS and nnetar are all provided as automatic forecasting techniques through the forecast package within the R software.
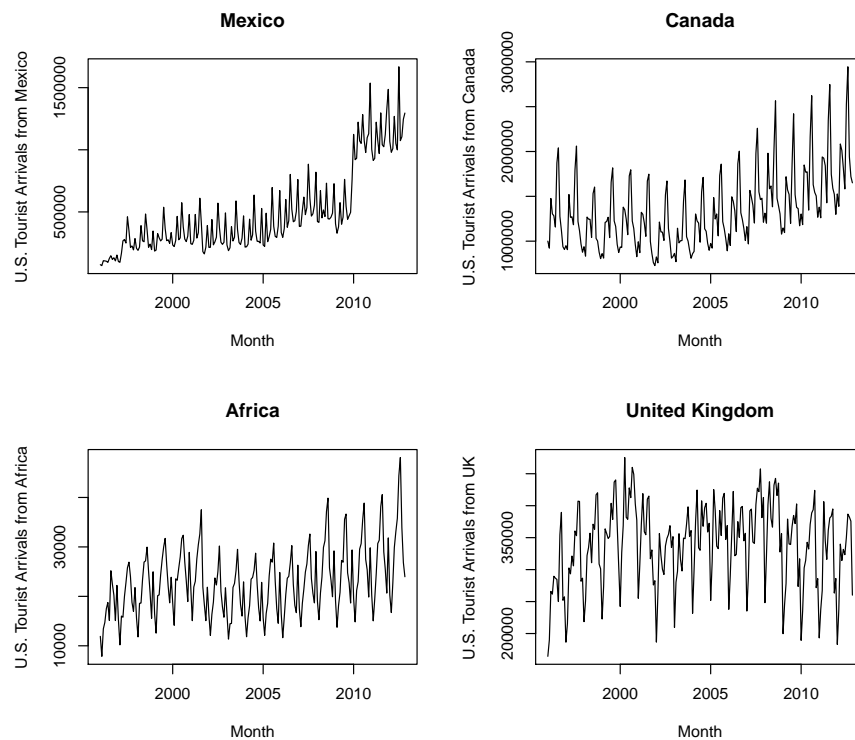
Fig. 5.4 Selected U.S. tourist arrivals time series (Jan. 1996 - Nov. 2012).

The results showed that the proposed SSA model of U.S. tourist arrivals out-performs all three of these models (ARIMA, ETS and Neural Networks). The *w*-correlations also provide an explanation for one reason behind the outstanding performance recorded by the SSA-V model as they clearly indicate that the SSA-V forecasting algorithm is highly successful in separating the signal from the noise found in the U.S. tourist arrivals series.

This application further uncovers substantial evidence to support the discontinuation of the use of ETS and feed-forward Neural Networks as models for forecasting inbound U.S. tourist arrivals in the future. This evidence is based on the MAPE, RMSE, RRMSE and DC criteria with statistically significant results. The results also show that the basic SSA-V model with optimal decomposition is able to outperform Automatic-ARIMA, ETS and nnetar models of Hyndman

and Khandakar (2008) in forecasting U.S. tourist arrivals. The results show that the nonparametric SSA-V model is on average 48% more accurate than the parametric model of ARIMA, 56% more accurate than ETS, and 68% more accurate than the feed-forward Neural Network model (nnetar) at forecasting tourist arrivals based on the RRMSE. In terms of the MAPE, SSA reports the lowest average MAPE at 8% in comparison to the ARIMA model's 14%, ETS' 17% and NN's 22% MAPEs. It is also noteworthy that SSA is the only model which is able to report a MAPE of less than 10% across all forecasting horizons which covers both the long and short run. This provides sound evidence for National Statistical Agencies in U.S. and elsewhere to consider introducing SSA as a more reliable method of forecasting tourist arrivals.

This application contributes to the literature on forecasting tourism demand in several ways. Firstly, it shows that the SSA technique can be used as a reliable demand forecasting technique for tourism in the future, using its application to inbound tourist arrivals in the U.S. as an example. This also results in an increase in the number of options available for demand forecasting in tourism. Secondly, the results show that SSA outperforms the ARIMA model of Hyndman and Khandakar (2008). This is an important finding as ARIMA models are widely used in forecasting tourism demand at present. Given the introduction of SSA and its strong performance with U.S. data it would be interesting to see how well the model performs in forecasting tourism demand in other nations. Thirdly, also evaluated is the performance of the SSA technique against an exponential smoothing and neural network model which shows the basic SSA-V forecasts are superior. The results are statistically significant and provide strong evidence to support the discontinuation of models such as ETS and feed forward Neural Networks with one hidden layer as tourism demand forecasting techniques for the U.S. Whilst more research work should be conducted on the comparison of SSA especially against neural networks in the future, the initial evidence is

supportive of the use of SSA.

Overall, given the importance of forecasting tourism demand and the important requirement that such forecasts be able to cope well with seasonality in demand, this application offers a new technique to forecasters in this area. The evidence from the U.S. data is that it offers the prospect of better forecasting accuracy than the pick of those techniques previously employed. Improvements in forecasting accuracy should provide a basis for more efficient resource allocation by, in particular, investors and managers in tourism.

In terms of the implications of this paper for further research there are several. Compared in this application is the performance of SSA to three of the most important existing alternative techniques. It would be worthwhile extending this analysis in the future to a wider range of alternative techniques. The encouraging results from employing SSA to forecast U.S. inbound tourism reported in this application also suggests that it may be worthwhile in future research to build a multivariate SSA model to forecast tourist arrivals. Here, it would be interesting to evaluate the spatio-temporal correlations between tourist arrivals from various countries (as proposed in Sato (2012)) so that this information could be used to enhance the multivariate SSA model to enable more accurate forecasts. Finally, the use of hybrid models has been common in the literature concerning the forecasting of tourism demand. It would be both interesting and of potential value for future research to consider how the SSA technique performs in a hybrid model. Moreover, the presence of structural breaks in U.S. tourist arrivals suggests that it would also be interesting to evaluate the impact on the forecasts of replacing KPSS tests with the Bai and Perron (2003) test for determining the differencing in ARIMA models. The results from forecasting tourist arrivals from Mexico also makes it clear that future studies should consider incorporating SSA change point detection for forecasting U.S. tourist arrivals. Finally, it is possible that different categories of tourism may be behaviourally different

in a way that is relevant for other forecasting uses. In this application, owing to data limitations it was not possible to analyse U.S. tourist arrivals based on purpose of visit and future research could benefit immensely if such data were made freely accessible and available by the relevant authorities.

Chapter 6

Conclusions

This thesis begins with an overall aim of ensuring lucrative theoretical developments in SSA which are well supported by empirical evidence. Chapter One begins with a general introduction which is followed by the introduction of the methodology in Chapter 2 in addition to the presentation of benchmark forecasting techniques and metrics used throughout this research. The remaining chapters (up until Chapter Seven) are organized such that each Chapter focuses on addressing the four objectives of this study. Accordingly, there are several contributions to the field.

The first contribution of this research was the introduction of a statistical test for comparing between the predictive accuracy of forecasts as presented via Chapter Three. This statistical test which exploits the concepts of stochastic dominance and cumulative distribution functions, extends the use of the Kolomogorov-Smirnov test and introduces it to the field of time series analysis and forecasting. In the past the application of the KS test was limited to its intended use which was the comparison of distributions between two data sets (Kolmogorov, 1933). This research has shown via both simulation studies and empirical results that the KS test can be extended as a method for comparing

between the predictive accuracy of forecasts to determine the existence of statistically significant differences between two sets of forecasts. In addition this alternate statistical test is able to overcome several problems with the original DM test (Diebold and Mariano, 1995) and the modified DM test (Harvey et al., 1997) pertaining to size, theoretical and applicability related issues. The simulation study was able to show that the proposed KSPA test is both better sized and more powerful than the modified DM test in Harvey et al. (1997), whilst the real world applications were used as evidence to illustrate the usefulness and applicability of the proposed test. This chapter opens up an entirely new area of research pertaining to the extension and improvement of the KSPA test. In particular, future studies should consider researching into the possibility of extending the test such that it could be used to compare between more than two forecasts at the same time and thereby further increase its practical value. In addition, the performance of the KSPA test in relation to other alternatives such as Hansen's Hansen (2005) Superior Predictive Ability (SPA) test and Hansen et al.'s Hansen et al. (2011) Model Confidence Set (MCS) should be evaluated.

The second contribution was the development of a new MSSA based theory for exploiting the forecastability of forecasts as presented via Chapter Four. This theoretical development which considers data with a time lag into the future, i.e. forecasts generated by official or professional forecasters for constructing a new and improved forecast, has shown that MSSA has the potential to extract auxiliary information from an existing forecast and create a more accurate forecast. An application to real data has illustrated the feasibility of the proposed theory. Forecast combining has been a long existing field of research in time series analysis. However, it is noteworthy that the proposed theoretical approach does not relate to this particular area where researchers seek to combine competing forecasts and develop a new forecast by exploiting variance-covariance based methods or regression based methods (see for example, Diebold and Lopez (1996)).

Instead, this novel approach considers combining historical data with a forecast which represents data with a time lag into the future and seeks to extract the auxiliary information contained within the forecast via multivariate modelling to develop a new and improved forecast. As such, having opened up an entirely new research avenue in the field of MSSA and multivariate time series analysis literature, future research should consider using this development alongside SSA change point detection and with automated algorithms which will promote the effective use of this new theoretical development. For example, firstly, an automated algorithm should be developed for extracting the VMSSA parameters for a given data set. Thereafter, extensive simulation studies which takes into account different noise levels, stationarity and non-stationarity amongst other time series features should be carried out to provide more justification for this theory.

The third and final contribution of this research (as discussed in Chapter Five) was the development of a new approach for the selection of parameters in SSA and MSSA. This was achieved by introducing new algorithms which enable both automation and optimization of SSA and MSSA forecasting by minimizing a loss function. This is a vital contribution to SSA as it opens up the possibility of using both univariate and multivariate versions for online forecasting applications and also enables users who are not conversant with the theory underlying SSA to continue exploiting this technique. This new automation of SSA and MSSA for forecasting contributes to existing literature and extends previous studies in several ways. Historically, the selection of SSA and MSSA parameters has been a highly labour intensive approach, see for example Elsner and Tsonis (1996); Golyandina et al. (2001); Hassani (2007). However, the approach presented in this thesis enables users to obtain the best values for L and r for out-of-sample forecasting without the need to analyse the data, periodogram or scree plot. In addition, the approach presented here has provided a form of statistical foundation for the selection of r by minimising a loss function. Such

statistical justification was lacking in the earlier approaches which were seen as highly subjective according to Khan and Poskitt (2013b). Moreover, the previous attempts at automating SSA focussed on trend and periodicity extraction (Alexandrov, 2009; Vautard et al., 1992), and denoising with SSA when faced with big signal to noise ratios (Alonso et al., 2004). This thesis extends this line of research by presenting a new approach for automating SSA and MSSA to obtain the best possible out-of-sample forecast. It should be noted that the approach presented in this thesis is not optimized for signal extraction, and instead is purely designed with a focus on forecasting. It is expected that this algorithm will result in an increased application of SSA in future as seen with models such as ARIMA, ETS and NN which are provided via the forecast package in R (Hyndman and Athanasopoulos, 2012). The application which follows after the introduction of automated SSA also has several contributions. Firstly, it shows that the automated SSA forecasting algorithm is able to outperform the automated ARIMA, ETS and NN algorithms by (Hyndman and Athanasopoulos, 2012) with statistically significant outcomes in relation to forecasting U.S. tourist arrivals. Secondly, the application also marks the first instance in which SSA is successfully applied for tourism demand forecasting and thereby adds to the extensive literature on forecasting tourist arrivals. Thirdly, the application rules out the use of ETS and a feed-forward NN model with one hidden layer for forecasting U.S. tourist arrivals in future and presents SSA as a viable alternative. Given the vast range of forecasting techniques available this study helps researchers to rule out two models whilst the need to compare SSA with several other models remains open. However, the algorithms presented herewith do not cover VMSSA. As such, future research should consider extending the proposed algorithms for VMSSA. In addition, at present these automated algorithms do not consider change point detection which can enable further improvements in forecast accuracy. As such, future research should also consider incorporating

SSA change point detection along with these automated algorithms.

In summary, this research has presented several important contributions to the field of time series analysis and SSA in particular. These range from the opening up of new research avenues via the introduction of an alternative to the Diebold-Mariano test for comparing between the predictive accuracy of two forecasts (Chapter 3) and a theoretical framework for exploiting the forecastability of forecasts (Chapter 4) to improving the user friendliness of a complex method such as SSA and MSSA (Chapter 5). In general, future research should consider comparing the proposed SSA and MSSA methods in relation to a variety of other benchmark techniques not included in this current work. These could include various other benchmark models provided through the forecast package (e.g. TBATS, ARFIMA) and other time series analysis and forecasting techniques such as GARCH and HAR.

References

Alexandrov, T. (2009). A method of trend extraction using singular spectrum analysis. *REVSTAT*, 7(1):1–22.

Alonso, J. F., Castillo, J. M. D., and Pintado, P. (2004). An automatic filtering procedure for processing biomechanical kinematic signals. In Barreiro, J. M., Martin-Sanchez, F., Maojo, V., and Sanz, F., editors, *Biological and Medical Data Analysis*. Springer, Berlin Heidelberg.

Álvarez-Díaz, M. and Rosselló-Nadal, J. (2010). Forecasting british tourist arrivals in the balearic islands using meteorological variables. *Tourism Economics*, 16(1):153–168.

Andrawis, R. R., Atiya, A. F., and El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, 27(3):870–886.

Armstrong, J. S. and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1):69–80.

Ash, J. C. K., Smyth, D. J., and Heravi, S. (1997). The accuracy of oecd forecasts for japan. *Pacific Economic Review*, 2(1):25–44.

Assaf, A. G., Barros, C. P., and Gil-Alana, L. A. (2011). Persistence in the short and long-term tourist arrivals to australia. *Journal of Travel Research*, 50(2):213–229.

Athanasopoulos, G. and de Silva, A. (2012). Multivariate exponential smoothing for forecasting tourist arrivals. *Journal of Travel Research*, 51(5):640–652.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104.

Beneki, C., Eeckels, B., and Leon, C. (2012). Signal extraction and forecasting of the uk tourism income time series: A singular spectrum analysis approach. *Journal of Forecasting*, 31(5):391–400.

Birnbaum, Z. W. and Tingey, F. H. (1951). One-sided confidence contours for probability distribution functions. *The Annals of Mathematical Statistics*, 22(4):592–596.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer.

Broomhead, D. S. and King, G. P. (1986a). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2):217–236.

Broomhead, D. S. and King, G. P. (1986b). On the qualitative analysis of experimental dynamical systems. *In: Nonlinear Phenomena and Chaos*, pages 113–144.

Cang, S. (2011). A non-linear tourism demand forecast combination model. *Tourism Economics*, 17(1):5–20.

Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7:1247–1250.

Chan, F. and Lim, C. (2011). Spectral analysis of seasonality in tourism demand. *Mathematics and Computers in Simulation*, 81(7):1409–1418.

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

Christiano, L. J. (1989). p*: not the inflation forecasters holy grail. *Federal Reserve Bank of Minneapolis Quarterly Review*, 13(4):3–18.

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

Chu, F. L. (2011). A piecewise linear approach to modeling and forecasting demand for macau tourism. *Tourism Management*, 32(6):1414–1420.

Clark, T. E. and McCracken, M. W. (2009). In-sample tests of predictive ability: A new approach. *Research Division Federal Reserve Bank of St. Louis Working Paper Series*, 2009-051A:1–34.

Clark, T. E. and McCracken, M. W. (2012). Nested forecast model comparisons: A new approach to testing equal accuracy. *Research Division Federal Reserve Bank of St. Louis Working Paper Series*, 2009-050B:1–59.

Clements, M. P. and Smith, J. (1999). A monte carlo investigation of forecasting performance of emprical setar model. *Journal of Applied Econometrics*, 14:123–141.

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

Coshall, J. T. and Charlesworth, R. (2011). A management orientated approach to combination forecasting of tourism demand. *Tourism Management*, 32(4):759–769.

Danilov, D. and Zhigljavsky, A. (1997). *Principal Components of Time Series: the 'Caterpillar' method*. University of St. Petersburg, St. Petersburg.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley.

Diebold, F. X. (2013). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. *Manuscript, Department of Economics, University of Pennsylvania*, pages 1–22.

Diebold, F. X. and Lopez, J. A. (1996). 8 forecast evaluation and combination. *Handbook of Statistics - Statistical Methods in Finance*, 14:241–268.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Elliot, G. and Timmermann, A. (2013). *Handbook of Economic Forecasting*. North Holland.

Elsner, J. B. and Tsonis, A. A. (1996). *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. Plenum.

Fourie, J. and Santana-Gallego, M. (2011). The impact of mega-sport events on tourist arrivals. *Tourism Management*, 32(6):1364–1370.

Fraedrich, K. (1986). Estimating the dimension of weather and climate attractors. *Journal of the Atmospheric Sciences*, 43(5):419–432.

Ghodsi, M., Hassani, H., Sanei, S., and Hicks, Y. (2009). The use of noise information for detection of temporomandibular disorder. *Biomedical Signal Processing and Control*, 4(2):79–85.

Ghodsi, Z., Silva, E. S., and Hassani, H. (2015). Bicoid signal extraction with a selection of parametric and nonparametric signal processing techniques. *Genomics, Proteomics & Bioinformatics*, 13(3):183–191.

Gil-Alana, L. A. (2005). Modelling international monthly arrivals using seasonal univariate long-memory processes. *Tourism Management*, 26(6):867–878.

Gilleland, E. and Roux, G. (2015). A new approach to testing forecast predictive accuracy. *Meteorological Applications*, 22(3):534–543.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

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

Golyandina, N. (2010). On the choice of parameters in singular spectrum analysis and related subspace-based methods. *Statistics and Its Interface*, 3(3):259–279.

Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of time series structure: SSA and related techniques*. CRC Press.

Gounopoulos, D., Petmezas, D., and Santamaria, D. (2012). Forecasting tourist arrivals in greece and the impact of macroeconomic shocks from the countries of tourists' origin. *Annals of Tourism Research*, 39(2):641–666.

Granger, C. W. J. and Newbold, P. (1977). *Forecasting Economic Time Series*. Academic Press, New York.

Greenidge, K. (2001). Forecasting tourism demand: An stm approach. *Annals of Tourism Research*, 28(1):98–112.

Groth, A. and Ghil, M. (2011). Multivariate singular spectrum analysis and the road to phase synchronization. *Physical Review E*, 84(3):036206.

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

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23(4):365–380.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). Model confidence set. *Journal of Business and Economic Statistics*, 79(2):453–497.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. *Journal of Data Science*, 5(2):239–257.

Hassani, H. (2010). A note on the sum of the sample autocorrelation function. *Physica A: Statistical Mechanics and its Applications*, 389(8):1601–1606.

Hassani, H., Dionisio, A., and Ghodsi, M. (2010a). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. *Nonlinear Analysis: Real World Applications*, 11(1):492–502.

Hassani, H., Heravi, S., Brown, G., and Ayoubkhani, D. (2013a). Forecasting before, during, and after recession with singular spectrum analysis. *Journal of Applied Statistics*, 40(10):2290–2302.

Hassani, H., Heravi, S., and Zhigljavsky, A. (2009). Forecasting european industrial production with singular spectrum analysis. *International journal of forecasting*, 25(1):103–118.

Hassani, H., Heravi, S., and Zhigljavsky, A. (2013b). Forecasting uk industrial production with multivariate singular spectrum analysis. *Journal of Forecasting*, 32(5):395–408.

Hassani, H., Leonenko, N., and Patterson, K. (2012a). The sample autocorrelation function and the detection of long-memory processes. *Physica A: Statistical Mechanics and its Applications*, 391(24):6367–6379.

Hassani, H. and Mahmoudvand, R. (2013). Multivariate singular spectrum analysis: A general view and new vector forecasting approach. *International Journal of Energy and Statistics*, 1(1):55–83.

Hassani, H., Mahmoudvand, R., Omer, H. N., and Silva, E. S. (2014). A preliminary investigation into the effect of outlier(s) on singular spectrum analysis. *Fluctuation and Noise Letters*, 13(4):1450029.

Hassani, H., Mahmoudvand, R., and Yarmohammadi, M. (2010b). Filtering and denoising in the linear regression model. *Fluctuation and Noise Letters*, 9(4):343–358.

Hassani, H., Mahmoudvand, R., and Zokaei, M. (2011). Separability and window length in singular spectrum analysis. *Comptes Rendus Mathematique*, 349(17-18):987–990.

Hassani, H., Mahmoudvand, R., Zokaei, M., and Ghodsi, M. (2012b). On the separability between signal and noise in singular spectrum analysis. *Fluctuation and Noise Letters*, 11(2):1250014.

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

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

Hassani, H. and Thomakos, D. (2010). A review on singular spectrum analysis for economic and financial time series. *Statistics and its Interface*, 3(3):377–397.

Heravi, S., Osborn, D. R., and Birchenhall, C. R. (2004). Linear versus neural network forecasts for european industrial production series. *International Journal of Forecasting*, 20:435–446.

Holt, C. C. (2004). Forecasting trends and seasonals by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10.

Horváth, L., Kokoszka, P., and Zitikis, R. (2006). Testing for stochastic dominance using the weighted mcfadden-type statistic. *Journal of Econometrics*, 133(1):191–205.

Hui, T. K. and Yuen, C. C. (2002). A study in the seasonal variation of japanese tourist arrivals in singapore. *Tourism Management*, 23(2):127–131.

Hyndman, R. J. and Athanasopoulos, G. (2012). *Forecasting: principles and practice*. O Texts.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *International Economics*, 27(3):1–22.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454.

Jackman, M. and Greenidge, K. (2010). Modelling and forecasting tourist flows to barbados using structural time series models. *Tourism and Hospitality Research*, 10(1):1–13.

Kapl, M. and Müller, W. G. (2010). Prediction of steel prices: A comparison between a conventional regression model and mssa. *Statistics and Its Interface*, 3:369–275.

Khan, M. A. R. and Poskitt, D. (2013a). Moment tests for window length selection in singular spectrum analysis of short– and long–memory processes. *Journal of Time Series Analysis*, 34(2):141–155.

Khan, M. A. R. and Poskitt, D. (2013b). A note on window length selection in singular spectrum analysis. *Australian and New Zealand Journal of Statistics*, 55(2):87–108.

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

Kim, S. S. and Wong, K. K. (2006). Effects of news shock on inbound tourist demand volatility in korea. *Journal of Travel Research*, 44(4):457–466.

Kolmogorov, A. N. (1933). Sulla determinazione emperica delle leggi di probabilita. *Giornale dell' Istituto Italiano degli Attuari*, 4:83–91.

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

Makridakis, S. G., Wheelright, S. C., and Hyndman, R. J. (1998). *Forecasting: Methods and Applications*. Wiley, New York.

Mao, C. K., Ding, C. G., and Lee, H. Y. (2010). Post-sars tourist arrival recovery patterns: An analysis based on a catastrophe theory. *Tourism Management*, 31(6):855–861.

Marsaglia, G., Tsang, W. W., and Wang, J. (2003). Evaluating kolmogorov's distribution. *Journal of Statistical Software*, 8(18):1–4.

McFadden, D. (1989). Testing for stochastic dominance. In Fomby, T. B. and Seo, T. K., editors, *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*. Springer, New York, Berlin, London, and Tokyo.

Meese, R. and Rogoff, K. (1988). Was it real? the exchange rate-interest rate differential relation over the modern floating-rate period. *Journal of Finance*, 43(4):933–948.

Morgan, W. A. (1939). A test for significance of the difference between two variances in a sample from a normal bivariate population. *Biometrika*, 31(1-2):13–19.

Naudé, W. A. and Saayman, A. (2005). Determinants of tourist arrivals in africa: a panel data regression analysis. *Tourism Economics*, 11(3):365–391.

Nelson, L. A., Dickey, D. A., and Smith, J. M. (2011). Estimating time series and cross section tourism demand models: Mainland united states to hawaii data. *Tourism Management*, 32(1):28–38.

Oropeza, V. and Sacchi, M. (2011). Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics*, 76(3):V25–V32.

Page, S., Song, H., and Wu, D. C. (2012). Assessing the impacts of the global economic crisis and swine flu on inbound tourism demand in the united kingdom. *Journal of Travel Research*, 51(2):142–153.

Patterson, K., Hassani, H., Heravi, S., and Zhigljavsky, A. (2011). Multivariate singular spectrum analysis for forecasting revisions to real-time data. *Journal of Applied Statistics*, 38(10):2183–2211.

Rodrıguez-Aragón, L. J. and Zhigljavsky, A. (2010). Singular spectrum analysis for image processing. *Statistics and Its Interface*, 3(3):419–426.

Sanei, S., Ghodsi, M., and Hassani, H. (2011). An adaptive singular spectrum analysis approach to murmur detection from heart sounds. *Medical Engineering & Physics*, 33(3):362–367.

Sanei, S. and Hassani, H. (2015). *Singular Spectrum Analysis of Biomedical Signals*. CRC Press.

Sato, A. H. (2012). Impact of the great east japan earthquake on hotel industry in pacific tohoku prefectures. *Progress of Theoretical Physics Supplement*, 194:165–172.

Seetanah, B., Durbarry, R., and Ragodoo, J. F. (2010). Using the panel cointegration approach to analyse the determinants of tourism demand in south africa. *Tourism Economics*, 16(3):715–729.

Seetaram, N. (2010). Use of dynamic panel cointegration approach to model international arrivals to australia. *Journal of Travel Research*, 49(4):414–422.

Shareef, R. and McAleer, M. (2007). Modelling the uncertainty in monthly international tourist arrivals to the maldives. *Tourism Management*, 28(1):23–45.

Shen, S., Li, G., and Song, H. (2008). An assessment of combining tourism demand forecasts over different time horizons. *Journal of Travel Research*, 47(2):197–207.

Shen, S., Li, G., and Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, 38(1):72–89.

Silva, E. S. and Hassani, H. (2015). On the use of singular spectrum analysis for forecasting u.s. trade before, during and after the 2008 recession. *International Economics*, 141:34–49.

Simard, R. and L'Ecuyer, P. (2011). Computing the two-sided kolmogorov-smirnov distribution. *Journal of Statistical Software*, 39(11):1–18.

Smeral, E. (2010). Impacts of the world recession and economic crisis on tourism: Forecasts and potential risks. *Journal of Travel Research*, 49(1):31–38.

Song, H., Dwyer, L., Li, G., and Cao, Z. (2012). Tourism economics research: A review and assessment. *Annals of Tourism Research*, 39(3):1653–1682.

Song, H., Gao, B. Z., and Lin, V. S. (2013). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, 29(2):295–310.

Song, H. and Li, G. (2008). Tourism demand modelling and forecasting - a review of recent research. *Tourism Management*, 29(2):203–220.

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

Tsay, R. S. (1998). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7(1):1–20.

Vautard, R. and Ghil, M. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D: Nonlinear Phenomena*, 35(3):395–424.

Vautard, R., Yiou, P., and Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signal. *Physica D: Nonlinear Phenomena*, 58(1-4):95–126.

Wan, S. K., Wang, S. H., and Woo, C. K. (2013). Aggregate vs. disaggregate forecast: Case of hong kong. *Annals of Tourism Research*, 42:434–438.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14(1):35–62.

Appendix A

R Codes

The following reports as examples, selected components of the R codes developed for the applications used in this study. The full code is available upon request.

Chapter 3

```
The Kolmogorov-Smirnov Predictive Accuracy test
# Install and load the "stats" package in R.
install.packages("stats")
library(stats)
# Input the forecast errors from two models. Let Error1 show
errors from the model with the lower error based on some
loss function.
Error1<-scan()
Error2<-scan()
# Convert the raw forecast errors into absolute values or
squared values
depending on the loss function.
```

```
abs1<-abs(Error1)
abs2<-abs(Error2)
sqe1<-(Error1)^2
sqe2<-(Error2)^2
# Perform the KSPA test for distinguishing between
the predictive accuracy of forecasts from the two models*.
#Two-sided KSPA test:
ks.test(abs1,abs2)
#One-sided KSPA test:
ks.test(abs1,abs2, alternative = c("greater"))
```

Chapter 5

Shown initially is a section of the code used to minimize the loss function in the SSA algorithms.

```
Forecast<-function(L,steps,M,S){
N=length(S);Lsize=length(L)
Forecast=array(dim=M)
MSE=array(0,dim=c((max(L)-1),Lsize))
for(l in 1:Lsize){
for(r1 in 1:(L[l]-1)){
for(i in 1:M){
X=S[1:(N-M-steps+i)]
if((N-M+i)<(N+1)){
Forecast[i]=VSSA.Forecasting(L[l],r1,X,steps)[steps]
MSE[r1,l]=MSE[r1,l]+(Forecast[i]-S[(N-M+i)])^2/(M)
}}}}
RMSE=sqrt(MSE)
```

```
dimnames(RMSE)=list(paste("r=",(1:(max(L)-1)),sep=""),
paste("L=",L,sep=""))
round( RMSE,4)
}
```

Presented next is a section of the code used to minimize the loss function in the
MSSA algorithms.

```
mg=ms=r1=r2=r3=r4=array(1000,dim=c(length(L),
length(L)))
#k1=M-L1+1;k2=M-L2+1;
N1=nrow(Y)-M-h+1
f1=f2=array(0,dim=c(N1,ncol(Y)))
for(j in 1:length(L)){
  r=seq(1,L[j]-1,1)
  for(i in 1:r[j]){
    for(p in 1:N1){
      ct=M+p-1
      f1[p,]= HMSSA.R(L[j],seq(1,r[i],1),h,Y[1:ct,])[h,]
      f2[p,]= HMSSA.V(L[j],seq(1,r[i],1),h,Y[1:ct,])[h,]
    }
    r1[i,j]=rmse(f1[,1],Y[(nrow(Y)-N1+1):nrow(Y),1])
    r2[i,j]=rmse(f1[,2],Y[(nrow(Y)-N1+1):nrow(Y),2])
    r3[i,j]=rmse(f2[,1],Y[(nrow(Y)-N1+1):nrow(Y),1])
    r4[i,j]=rmse(f2[,2],Y[(nrow(Y)-N1+1):nrow(Y),2])
  }
}
rl1=which(r1 == min(r1), arr.ind = TRUE)
rl2=which(r2 == min(r2), arr.ind = TRUE)
```

```
rl3=which(r3 == min(r3), arr.ind = TRUE)

rl4=which(r4 == min(r4), arr.ind = TRUE)

min(r1)

min(r2)

min(r3)

min(r4)
```