Development of the theoretical and methodological aspects of the singular spectrum analysis and its application for analysis and forecasting of economics data

Hossein Hassani

Cardiff School of Mathematics, Cardiff University

December 2009

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date 9-Oct-2009

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Signed ..... (candidate) Date 9-Oct-2009

STATEMENT 2

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit reference. A bibliography is appended.

Signed ..... (candidate) Date 9-Oct-2009

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed ..... (candidate) Date 9-Oct-2009

Abstract

In recent years Singular Spectrum Analysis (SSA), used as a powerful technique in time series analysis, has been developed and applied to many practical problems. The aim of this research is to develop theoretical and methodological aspects of the SSA technique and to demonstrate that SSA can be considered as a powerful method of time series analysis and forecasting, particulary for economic time series.

For practical aspect and empirical results, various economic and financial time series are used. First, the SSA technique is applied as a noise reduction method. The performance of SSA is examined in noise reduction of several important financial series. The daily closing prices of several stock market indices are examined to analyse whether noise reduction matters in measuring dependencies of the financial series. The effect of noise reduction is considered on the linear and nonlinear measures of dependence between two series. The results are compared with those obtained with the linear and nonlinear methods for filtering time series. The results show that the performance of SSA is much better than of the competitive methods.

Second, we consider the performance of SSA in forecasting various time series. For consistency with the forecasting results obtained with other current forecasting methods, the performance of the SSA technique is examined by applying it to a well-known time series data set, namely, monthly accidental deaths in the USA. The results are compared with those obtained using Box-Jenkins SARIMA models, the ARAR algorithm and the Holt-Winter algorithm. The results show

that the SSA technique gives a much more accurate forecast than the other methods indicated above.

As another example, the performance of the SSA technique is assessed by applying it to 24 series measuring the monthly seasonally unadjusted industrial production for important sectors of the German, French and UK economies. The results confirm that at longer horizons, SSA significantly outperforms ARIMA and Holt-Winter methods.

Moreover, the application of SSA to the analysis and forecasting of Iranian national accounts data, which are rather short, are considered to examine capability of SSA in forecasting short time series. The results confirm that SSA works very well for short time series as well as for long time series.

The univariate and multivariate SSA are also employed in predicting the value and the changes in direction of inflation series for the United States. The consumer price indices, and real-time chain-weighted GDP price index series are used in these prediction exercises. Moreover, our out-of-sample h-step-ahead moving prediction results are compared with the prediction results based on methods such as activity-based NAIRU Philips curve, $AR(p)$, and random walk models with the latter as a naive forecasting method. A short-run (quarterly) and long-run (one to six years) time windows are utilized for predictions. The results clearly confirm that prediction of inflation rate in the United States during the period of "Great Moderation" is less challenging compared to more volatile inflationary period of 1970-1985 also.

Furthermore, the univariate and multivariate SSA is used for predicting the value and the direction of changes in the daily pound/dollar exchange rate. Empirical results show that the forecast based on the

multivariate SSA compares favorably to the forecast of the random walk model both for predicting the value and the direction of changes in the daily pound/dollar exchange rate. The SSA forecasting results are also compared to prediction results based on an error correction model (VEC) in the context of a restricted vector autoregressive model. The results show that the VEC results are inferior.

For theoretical development of the technique, two new versions of SSA are introduced; the SSA technique based on the minimum variance estimator and based on the perturbation theory. The new versions are examined in reconstructing and forecasting time series. The results are compared with the current version of SSA and indicate that the new versions improve the quality of reconstruction step as well as forecasting results.

We also consider the concept of casual relationship between two time series based on the SSA technique. We introduce several criteria which characterize this causality. The criteria are based on the forecasting accuracy and predictability of the direction of change. The performance of the proposed test is examined using different real time series.

Contents

Acknowledgements

I would like to thank my supervisor Professor A. Zhigljavsky, for his invaluable support and guidance throughout this PhD. The creativity and enthusiasm of him for Singular Spectrum Analysis has inspired me to work on this topic.

I am indebted to Dr. V. Netrutkin and Dr. N. Golyandina from St. Petersburg University, who have contributed significantly to the provision of materials required for this research.

I would like to special thanks Central Bank of the Islamic Republic of Iran for the financial support they have provided throughout this research.

I would also like to thank my external examiner, Prof. C. Ioannidis, and my internal examiner, Dr V. Moskvina, for taking the time to read my thesis and for their useful suggestions that led to the improvement of this thesis.

My appreciation and gratefulness to the friendly staff of the Cardiff School of Mathematics for their help and kind at various times during my research.

I dedicate this thesis to my wife

MANSI

who did more than her fair share in life and love while I worked on this thesis. Mansi is my source of strength and without her patient love and her support this work would never have started much less finished.

Publications

1. Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5(2), pp. 239–257.

2. Hassani, H; Heravi, H; and Zhigljavsky, A. (2009). Forecasting European Industrial Production with Singular Spectrum Analysis, *International journal of forecasting*, 25(1), pp. 103–118.

3. Hassani, H; and Zhigljavsky, A. (2009). Singular Spectrum Analysis: Methodology and Application to Economics Data, *Journal of Systems Science and Complexity (JSSC)*, 22(3), pp. 372–394.

4. Hassani, H. (2010). Singular Spectrum Analysis Based on the Minimum Variance Estimator, *Nonlinear Analysis: Real World Applications*, Forthcoming.

5. Hassani, H; Soofi, A; and Zhigljavsky, A. (2010). Predicting Daily Exchange Rate with Singular Spectrum Analysis, *Nonlinear Analysis: Real World Applications*, Forthcoming.

6. Hassani, H; Zokaei, M; von Rosen, D; Amiri, S; and Ghodsi, M. (2009). Does noise reduction matter for curve fitting in growth curve models?, *Computer Methods and Programs in Biomedicine*, 96(3), pp. 173–181.

7. Ghodsi, M; Hassani, H; Sanei, S; and Hicks, Y. (2009). The Use of Noise Information for Detection of Temporomandibular Disorder, *Journal of Biomedical Signal Processing and Control*, 4(2), pp. 79–85.

8. Hassani, H; Thomakos, D. (2010). A Review on Singular Spectrum Analysis for Economic and Financial Time Series, *Statistics and Its Interface*, Forthcoming.

9. Hassani, H; Zhigljavsky, A; Patterson, K; and Soofi, A. (2010). A Comprehensive Causality Test Based on the Singular Spectrum Analysis, *Causality in Science*, Oxford University press, Forthcoming.

10. Patterson, K; Hassani, H; Heravi, S; and Zhigljavsky, A. (2010). Forecasting the Final Vintage of the Index of Industrial Production, *Journal of Applied Statistics*, Forthcoming.

11. Ghodsi, M; Hassani, H; and Sanei, S. (2010). Extracting Fetal Heart Signal From Noisy Maternal ECG by Singular Spectrum Analysis, *Statistics and Its Interface*, Forthcoming.

12. Hassani, H., Zhigljavsky, A; and Xu, Z. (Submitted). Singular Spectrum Analysis Based on the Perturbation Theory.

13. Hassani, H; Soofi, A; and Zhigljavsky, A. (Submitted). Predicting Inflation Dynamics with Singular Spectrum Analysis.

14. Sanei, S.; Ghodsi, M; and Hassani, H. (Submitted). A constrained Singular Spectrum Analysis Approach to Murmur Detection from Heart Sounds.

List of Acronyms

SVD	Singular Value Decomposition
SSA	Singular Spectrum Analysis
LRF	Linear Recurrent Formula
CDF	Cumulative Distribution Function
RW	Random Walk
ACF	Autocorrelation Function
DFA	Detrended Fluctuation Analysis
DMA	Detrended Moving Average
ARMA	Autoregressive Moving Average
AR	Autoregressive
VAR	Vector Autoregressive
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
RMSE	Root Mean Square Error
CPI	Consumer Price Index

GDP	Gross Domestic Product
GNP	Gross National Product
EMH	Efficient Market Hypothesis

List of Symbols

$\|a\|$	The absolute value of a
$a.b$	The dot product of a and b
$\mathbf{A}^T$	The transpose of matrix $\mathbf{A}$
X^T	The transpose of vector X
$\|X\|$	The Euclidean norm of vector X
$\|\mathbf{A}\|_F$	The Frobenius norm of matrix $\mathbf{A}$
$\|\mathbf{A}\|_E$	The Euclidean norm of matrix $\mathbf{A}$
$\langle , \rangle$	The inner product
Y_Δ	The last $L-1$ components of the vector Y
Y^∇	The first $L-1$ components of the vector Y
$\mathfrak{L}_r$	Linear Space
$\mathcal{H}$	Hankelization procedure
$\mathcal{P}$	Linear operator

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Econometric methods have been widely used to forecast the evolution of quarterly and yearly national account data sets. For example, accurate prediction of inflation rate has been a subject of great research interest for economists. Accurate prediction of inflation plays an important role in macroeconomic policy analysis and decision making. However, many of the structural or time series forecasting models have failed to predict accurately economic time series.

On the other hand, many factors could affect the national economies and hence the national account data which are at best inaccurate representation of the macroeconomic variables because of measurement noise. The exogenous factors that cause instability in macroeconomics include technological changes, government policy changes, changes in the preferences of the consumers, and other events. These shocks cause structural changes in these time series making them nonstationary. Development of a methodology which is robust under these changes, is of paramount importance in accurate prediction of macroeconomic time series.

There are several reasons why classical model does not have a good performance for modelling and forecasting economic and financial series. First, an economic model that has been established to have validity

in explaining a relationship under one set of assumptions is useless if the assumptions are not valid. Model assumptions include not only those that can be expressed as predicates on model parameters but others with more qualitative or asymptotic form (for more information see [1]).

Moreover, many structural econometric and time series models devised for forecasting macroeconomic time series are based on restrictive assumptions of normality and linearity of the observed data. The methods that do not depend on these assumptions could be very useful for modelling and forecasting economics data. On the other hand classical methods of forecasting such as ARIMA type models are based on the assumption such as stationarity of the series and normality of residuals (see, for example, [2], [3] and references therein) .

Furthermore, it is well known that noise can seriously limit accuracy of time series prediction. Currently there are not many effective forecasting techniques available when there is significant noise in the time series data.

In general, there are two main approaches for forecasting noisy time series. According to the first one, we ignore the presence of noise and fit a forecasting model directly from noisy data hoping to extract the underlying deterministic dynamics. According to the second approach, which is often more effective than the first one, we start with filtering the noisy time series in order to reduce the noise level and then forecast the new data points (see, for example, [4, 5] and references therein). There are several linear and nonlinear noise reduction methods such as ARMA model, local projective, singular value decomposition (SVD) and simple nonlinear filtering. It is currently accepted that SVD-based

methods are very effective for the noise reduction in deterministic time series and correspondingly for forecasting [5].

Additionally, some of the previous research have considered economic and financial time series as deterministic, linear dynamical systems. In this case, the linear models can be used for modelling and forecasting. However, it has been shown that most of the financial time series are nonlinear (see, for example, [4–7]); in these cases, we should use nonlinear methods. Having a method that works well for both linear and nonlinear, stationary and non stationary time series is ideal for modelling and forecasting. The Singular Spectrum Analysis (SSA) meets all conditions stated above. The SSA technique is a non-parametric technique of time series analysis incorporating the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing [8]. Note also that SSA naturally incorporates the filtering of the series and the SVD.

The appearance of SSA is usually associated with the publication of papers by Broomhead and King [9] while the ideas of SSA were simultaneously developed in Russia (St. Petersburg, Moscow) and in several groups in the UK and USA [8, 11]. A thorough description of the theoretical and practical foundations of the SSA technique (with many examples) can be found in [8,10]. An elementary introduction to the subject can be found in [11]. Below we describe several applications of SSA and provide a brief discussion on the methodology used.

The basic SSA method consists of two complementary stages: decomposition and reconstruction; both stages include two separate steps. At the first stage we decompose the series and at the second stage we reconstruct the original series and use the reconstructed series for fore-

casting new data points. The main concept in studying the properties of SSA is 'separability', which characterizes how well different components can be separated from each other. The absence of approximate separability is often observed in series with complex structure. For these series and series with special structure, there are different ways of modifying SSA leading to different versions such as SSA with single and double centering, Toeplitz SSA, and sequential SSA [8].

On the other hand, asymptotic separation plays a very important role in the theory of SSA. It has been observed that in many practical applications the asymptotic features (which hold as the length of the series T tends to infinity) are met for relatively small values of T; it is not uncommon to successfully apply SSA to series with T equal to 20–30. Another important feature of SSA is that it can be used for analyzing relatively short series. I has been shown that SSA works very well for short time series as well as for long time series in forecasting macro-economics data [12].

It is worth noting that although some probabilistic and statistical concepts are employed in the SSA-based methods, we do not have to make any statistical assumptions such as stationarity of the series or normality of the residuals. Therefore, SSA is a very useful tool which can be used for solving the following problems:

finding trends of different resolution;

smoothing;

extraction of seasonality components;

simultaneous extraction of cycles with small and large periods;

extraction of periodicities with varying amplitudes;

simultaneous extraction of complex trends and periodicities;

finding structure in short time series.

Solving all these problems corresponds to the so-called basic capabilities of SSA. In addition, the method has several essential extensions. First, the multivariate version of the method permits the simultaneous expansion of several time series; see, for example [10]. Second, the SSA ideas lead to several forecasting procedures for time series; see [8, 10]. Also, the same ideas are used in [8] and [13] for change-point detection in time series. For comparison with classical methods, ARIMA, ARAR algorithm and Holt-Winter, see [14]– [16]. For automatic methods of identification within the SSA framework see [17] and for recent work in 'Caterpillar'-SSA software as well as new developments see [18].

Let us mention some other areas related to SSA. A variety of techniques of time series analysis and signal processing have been suggested that use SVD of certain matrices; for surveys see, for example, [19, 20]. Most of these techniques are based on the assumption that the original series is random and stationary; they include some techniques that are famous in signal processing, such as Karhunen-Loeve decomposition (for signal processing references see, for example [21]). Some statistical aspects of the SVD-based methodology for stationary series are considered, for example, in [22] and [23, 24].

The analysis of periodograms is an important part of the process of identifying the components in the SSA decomposition. A comparison of the observed spectrum of some common time series (these can be found, for example, in [25] and [26], Chapter 11) can help in understanding the nature of the residuals and in the formulation of the proper statistical hypothesis concerning the noise.

The idea of using dynamical systems theory for analyzing financial time series can be justified using the argument that the traditional statistical methods have only very limited success in real world financial applications; this is due to the fact that the financial time series have very complicated dynamical behaviour, see e.g. [4].

Another area which SSA is related to, is nonlinear (deterministic) time series analysis. It is a fashionable area of rapidly growing popularity; see, for example, recent books [27–30]. In the area of nonlinear time series analysis SSA was considered as a technique that could compete with more standard methods. There is a number of studies that considered SSA as a filtering method in (see, for example, [31] and references therein). The superiority of the SSA technique over traditional digital filtering methods used in biomedical data was shown, with several examples in the literature [32]. In another study, the noise information extracted using the SSA technique, has been used as a biomedical diagnostic test [33]. The SSA technique also used as a filtering method for longitudinal measurements. It has been shown that noise reduction is important for curve fitting in growth curve models, and that SSA can be employed as a powerful tool for noise reduction for longitudinal measurements [34].

Here we use the SSA technique for analysis, filtering, and forecasting financial and economic time series. The univariate and multivariate version of the SSA technique is used in this predictions which include both the magnitude and direction of changes.

The structure of this thesis is as follows. A brief introduction of the SSA method is represented in Chapter 2. In Chapter 3, we consider the SSA technique as a noise reduction method. The performance of

SSA as a forecasting method is considered in Chapter 4. Two new versions of SSA, SSA based on the minimum variance estimator and SSA based on the perturbation theory, are introduced in Chapters 5 and 6. A new casuality test based on the SSA technique is introduced in Chapter 7. Finally, Chapter 8 presents a summary of the study and some concluding remarks.

Chapter 2

SINGULAR SPECTRUM ANALYSIS

The main purpose of SSA is to decompose the original series into a sum of series, so that each component in this sum can be identified as either a trend, periodic or quasi-periodic component (perhaps, amplitude-modulated), or noise. This is followed by a reconstruction of the original series. The Basic SSA technique is performed in two stages, both of which include two separate steps as follows:

- Stage 1 : Decomposition
  - Step 1 : Embedding
  - Step 2 : Singular Value Decomposition (SVD)
- Stage 2 : Reconstruction
  - Step 1 : Grouping
  - Step 2 : Diagonal Averaging

A short description of the SSA technique is given as follows (for more information see [8]).

2.1 Decomposition

1st step: Embedding

Embedding can be regarded as a mapping that transfers a one-dimensional time series $Y_T = (y_1, \ldots, y_T)$ into the multidimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})^T \in \mathbf{R}^L$, where $K = T - L + 1$. Vectors X_i are called *L-lagged vectors* (or, simply, *lagged vectors*). The single parameter of the embedding is the *window length* L, an integer such that $2 \leq L \leq T$. The result of this step is the trajectory matrix

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

Note that the trajectory matrix $\mathbf{X}$ is a Hankel matrix, which means that all the elements along the diagonal $i + j = const$ are equal. Embedding is a standard procedure in time series analysis. With the embedding performed, future analysis depends on the aim of the investigation. For specialists in dynamical systems, a common technique is to obtain the empirical distribution of all pairwise distances between the lagged vectors X_i and X_j and then calculate the so-called correlation dimension of the series. Note that in this approach, L must be relatively small and K must be very large (formally, $K \to \infty$). The approximation of a stationary series with the help of the autoregression model can also

be expressed in terms of embedding: if we deal with the model

$$y_{i+L-1} = a_{L-1}y_{i+L-2} + \cdots + a_1 y_i + \varepsilon_{i+L-1}, \qquad i \geq 1$$

then we search for vector $A = (a_1, \ldots, a_{L-1}, -1)^T$ such that the scalar products (X_i, A) are described in terms of certain noise series.

2nd step: Singular Value Decomposition (SVD)

The second step, the SVD step, makes the singular value decomposition of the trajectory matrix $\mathbf{X}$ and represents it as a sum of rank-one bi-orthogonal elementary matrices. Denote by $\lambda_1, \ldots, \lambda_L$ the eigenvalues of $\mathbf{X}\mathbf{X}^T$ in decreasing order of magnitude ($\lambda_1 \geq \ldots \lambda_L \geq 0$) and by $U_1, \ldots, U_L$ the orthonormal system of the eigenvectors of the matrix $\mathbf{X}\mathbf{X}^T$ corresponding to these eigenvalues. Set

$$d = \max(i, \text{ such that } \lambda_i > 0) = rank\, \mathbf{X}.$$

If we denote $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, then the SVD of the trajectory matrix can be written as:

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d, \tag{2.1.1}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$. The matrices $\mathbf{X}_i$ have rank 1 (thus they are elementary matrices); U_i (in SSA literature they are called 'factor empirical orthogonal functions' or simply EOFs) and V_i (often called 'principal components') are the left and right eigenvectors of the trajectory matrix. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ is called the i-th eigentriple of the matrix $\mathbf{X}$, $\sqrt{\lambda_i}$ $(i = 1, \ldots, d)$ are the singular values of the matrix $\mathbf{X}$ and the set $\{\sqrt{\lambda_i}\}$ is called the spectrum of the matrix $\mathbf{X}$. If all eigenvalues

have multiplicity one, then the expansion (2.1.1) is uniquely defined.

SVD (2.1.1) is optimal in the sense that among all the matrices $\mathbf{X}^{(r)}$ of rank $r < d$, the matrix $\sum_{i=1}^{r} \mathbf{X}_i$ provides the best approximation to the trajectory matrix $\mathbf{X}$, so that $\| \mathbf{X} - \mathbf{X}^{(r)} \|$ is minimum. Here the norm of a matrix $\mathbf{Y}$ is defined as $\sqrt{\langle \mathbf{Y}, \mathbf{Y} \rangle}$, where the scalar product of two matrices $\mathbf{Y} = (y_{ij})_{i,j=1}^{q,s}$ and $\mathbf{Z} = (z_{ij})_{i,j=1}^{q,s}$ is $\langle \mathbf{Y}, \mathbf{Z} \rangle = \sum_{i,j=1}^{q,s} y_{ij} z_{ij}$. Note that $\| \mathbf{X} \|^2 = \sum_{i=1}^{d} \lambda_i$ and $\| X_i \|^2 = \lambda_i$ for $i = 1, \ldots, d$. Thus, we can consider the ratio $\lambda_i / \sum_{i=1}^{d} \lambda_i$ as the characteristic of the contribution of the matrix $\mathbf{X}_i$ to expansion (2.1.1). Consequently, $\sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{d} \lambda_i$, the sum of the first r ratios, is the characteristic of the optimal approximation of the trajectory matrix by the matrices of rank r.

Another optimal feature of the SVD is related to the properties of the directions determined by the eigenvectors $U_1, \ldots, U_d$. Specifically, the first eigenvector U_1 determines the direction such that the variation of the projections of the lagged vectors into this direction is maximum. Every subsequent eigenvector determines the direction that is orthogonal to all previous directions, and the variation of the projection of the lagged vectors onto this direction is also maximum. Therefore, it is natural to call the direction of the i-th eigenvector U_i the *i-th principal direction.* Note that the elementary matrices $\mathbf{X}_i$ are built up from the projections of the lagged vectors onto the i-th particular directions. This view on the SVD of the trajectory matrix composed of L-lagged vectors and an appeal to association with the *principal component analysis* lead to the following terminology. We shall call the vector U_i the i-th eigenvector, the vector V_i will be called the *i-th factor vector* and the vector $Z_i = \sqrt{\lambda_i} V_i$ the *i-th principal component.*

2.2 Reconstruction

1st Step: Grouping

The grouping step corresponds to splitting the elementary matrices into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_p\}$ be 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 $J = \{1, \ldots, d\}$ into disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation

$$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}. \tag{2.2.1}$$

The procedure of choosing the sets $I_1, \ldots, I_m$ is called the eigentriple grouping. For a given group I the contribution of the component $\mathbf{X}_I$ in the expansion (2.2.1) is measured by the share of the corresponding eigenvalues: $\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$. If the matrix $\mathbf{X}_I$ is a Hankel matrix, then there exist series $Y_T^{(1)}$ and $Y_T^{(2)}$ such that $Y_T = Y_T^{(1)} + Y_T^{(2)}$ and the trajectory matrices of these series are $\mathbf{X}_I$ and $\mathbf{X}_{J \setminus I}$, respectively. If the matrices $\mathbf{X}_I$ and $\mathbf{X}_{J \setminus I}$ are approximately Hankel matrices then the trajectory matrices of the series $Y_T^{(1)}$ and $Y_T^{(2)}$ are close to $\mathbf{X}_I$ and $\mathbf{X}_{J \setminus I}$. In this case we shall say that the series are approximately separable, see [8] for many more details. Therefore, the purpose of the grouping step (that is, the procedure of arranging the indices $1, \ldots, d$ into groups) is to find several groups $I_1, \ldots, I_m$ such that the matrices $\mathbf{X}_{I_1}, \ldots, \mathbf{X}_{I_m}$ satisfy (2.2.1) and are close to certain Hankel matrices. The grouping step is based on the analysis of the eigenvectors U_i and V_i, and eigenvalues λ_i in the SVD expansion. The principles and methods of identifying the SVD components for their inclusion into different

groups are described in [8], Sect. 1.6. Since each matrix component of the SVD is completely determined by the corresponding eigentriple, we shall talk about the grouping of the eigentriples rather than the grouping of the elementary matrices $\mathbf{X}_i$.

2nd Step: Diagonal averaging

The purpose of diagonal averaging is to transform a matrix to the form of a Hankel matrix which can be subsequently converted to a time series. If 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 $i + j = k + 1$. This procedure is called diagonal averaging, or Hankelization of the matrix $\mathbf{Z}$. The result of the Hankelization of a matrix $\mathbf{Z}$ is the Hankel matrix $\mathcal{H}\mathbf{Z}$. Note that the Hankelization is an optimal procedure in the sense that the matrix $\mathcal{H}\mathbf{Z}$ is the nearest to $\mathbf{Z}$ (with respect to the matrix norm) among all Hankel matrices of the corresponding size (see [8], Sect. 6.2). In its turn, the Hankel matrix $\mathcal{H}\mathbf{Z}$ uniquely defines the series by relating the value in the diagonals to the values in the series.

If z_{ij} stands for an element of a matrix $\mathbf{Z}$, then the k-th term of the resulting time series is obtained by averaging z_{ij} over all i, j such that $i + j = k + 2$. This procedure is called *diagonal averaging*, or Hankelization of the matrix $\mathbf{Z}$. The result of the Hankelization of a matrix $\mathbf{Z}$ is the Hankel matrix $\mathcal{H}\mathbf{Z}$, which is the trajectory matrix corresponding to the time series obtained as a result of the diagonal averaging.

The operator $\mathcal{H}$ acts on an arbitrary $L \times K$-matrix $\mathbf{Z} = (z_{ij})$ with

$L \leq K$ in the following way: for $i + j = s$ and $N = L + K - 1$ the element $\widetilde{z}_{ij}$ of the matrix $\mathcal{H}\mathbf{Z}$ is

$$\begin{cases} \dfrac{1}{s-1}\displaystyle\sum_{l=1}^{s-1} z_{l,s-l} & 2 \leq s \leq L-1, \\ \dfrac{1}{L}\displaystyle\sum_{l=1}^{L} z_{l,s-l} & L \leq s \leq K+1, \\ \dfrac{1}{K+L-s+1}\displaystyle\sum_{l=s-K}^{L} z_{l,s-l} & K+2 \leq s \leq K+L. \end{cases}$$

Note that the Hankelization is an optimal procedure in the sense that the matrix $\mathcal{H}\mathbf{Z}$ is the nearest to $\mathbf{Z}$ (with respect to the Frobenius norm) among all Hankel matrices of the corresponding size. Note that the Frobenius norm is equal to the square root of the matrix trace of $\mathbf{X}\mathbf{X}^T$. The Hankel matrix $\mathcal{H}\mathbf{Z}$ uniquely defines the time series by relating the values in the diagonals to the values in the series.

By applying the Hankelization procedure to all matrix components of (2.2.1), we obtain another expansion:

$$\mathbf{X} = \widetilde{\mathbf{X}}_{I_1} + \ldots + \widetilde{\mathbf{X}}_{I_m} \tag{2.2.2}$$

where $\widetilde{\mathbf{X}}_{I_1} = \mathcal{H}\mathbf{X}$. This is equivalent to the decomposition of the initial series $Y_T = (y_1, \ldots, y_T)$ into a sum of m series:

$$y_t = \sum_{k=1}^{m} \widetilde{y}_t^{(k)} \tag{2.2.3}$$

where $\widetilde{Y}_T^{(k)} = (\widetilde{y}_1^{(k)}, \ldots, \widetilde{y}_T^{(k)})$ corresponds to the matrix $\mathbf{X}_{I_k}$. A sensible grouping leads to the decomposition (2.1.1) where the resultant matrices $\mathbf{X}_{I_k}$ are almost Hankel ones. This corresponds to approximate separability and implies that pairwise scalar products of different ma-

trices $\mathbf{X}_{I_k}$ in (2.2.2) are small. The procedure of computing the time series $\widetilde{Y}_T^{(k)}$ (that is, building up the group I_k plus diagonal averaging of the matrix $\mathbf{X}_{I_k}$) will be called *reconstruction* of a series $Y_T^{(k)}$ by the eigentriples with indices in I_k. In relation to the grouping method, it is worthwhile to note that if L is large enough, the eigenvectors in a sense imitate the behavior of the corresponding time series components. In particular, the trend of the series corresponds to slowly varying eigenvectors. The harmonic component produces a pair of left (and right) harmonic eigenvectors with the same frequency, etc.

2.3 Reconstruction Algorithm

To formalize the SSA reconstruction step, let us have a time series $Y_T = (y_1, \ldots, y_T)$. Fix L ($L \leq T/2$), the window length, and let $K = T - L + 1$.

Step 1. (*Computing the trajectory matrix*): transfers a one-dimensional time series $Y_T = (y_1, \ldots, y_T)$ into the multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})' \in \mathbf{R}^L$, where $K = T - L + 1$. Vectors X_i are called *L-lagged vectors* (or, simply, *lagged vectors*). The single parameter of the embedding is the *window length* L, an integer such that $2 \leq L \leq T$. The result of this step is the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K] = (x_{ij})_{i,j=1}^{L,K}$.

Step 2. (*Constructing a matrix for applying SVD*): compute the matrix $\mathbf{X}\mathbf{X}^T$.

Step 3. (*SVD of the matrix* $\mathbf{X}\mathbf{X}^T$): compute the eigenvalues and eigen–vectors of the matrix $\mathbf{X}\mathbf{X}^T$ and represent it in the form $\mathbf{X}\mathbf{X}^T = P\Lambda P^T$. Here $\Lambda = diag(\lambda_1, \ldots, \lambda_L)$ is the diagonal matrix of

eigenvalues of $\mathbf{X}\mathbf{X}^T$ ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L \geq 0$ and $P = (P_1, P_2, \ldots, P_L)$ is the corresponding orthogonal matrix of eigen–vectors of $\mathbf{X}\mathbf{X}^T$.

Step 4. (*Selection of eigen–vectors*): select a group of l ($1 \leq l \leq L$) eigen–vectors $P_{i_1}, P_{i_2}, \ldots, P_{i_l}$.

The grouping step corresponds to splitting the elementary matrices $\mathbf{X}_i$ into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_l\}$ be a group of indices $i_1, \ldots, i_l$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_l}$.

Step 5. (*Reconstruction of the one-dimensional series*): compute the matrix $\tilde{\mathbf{X}} = ||\tilde{x}_{i,j}|| = \sum_{k=1}^{l} P_{i_k} P_{i_k}^T \mathbf{X}$ as an approximation to $\mathbf{X}$. Transition to the one–dimensional series can now be achieved by averaging over the diagonals of the matrix $\tilde{\mathbf{X}}$.

2.4 Forecasting Algorithm

Forecasting by SSA can be applied to the time series that approximately satisfy linear recurrent formulae (LRF):

$$y_{i+d} = \sum_{k=1}^{d} a_k y_{i+d-k}, \qquad 1 \leq i \leq T - d \tag{2.4.1}$$

of some dimension d with the coefficients $a_1, \ldots, a_d$. An important property of the SSA decomposition is that, if the original time series Y_T satisfies a LRF, then for any T and L there are at most d nonzero singular values in the SVD of the trajectory matrix $\mathbf{X}$; therefore, even if the window length L and $K = T - L + 1$ are larger than d, we only need at most d matrices $\mathbf{X}_i$ to reconstruct the series.

SSA forecasting algorithm is based on a premise which, roughly

speaking, states that: If the number of terms r in the SVD of the trajectory matrix $\mathbf{X}$ is smaller than the window length L, then the series satisfies some LRF of some dimension $d \leq r$. Let us formally describe the forecasting algorithm under consideration (for more information see [8]):

Algorithm input:

(a) Time series $Y_T = (y_1, \ldots, y_T)$.

(b) Window length L, $1 < L < T$.

(c) Linear Space $\mathfrak{L}_r \subset \mathbf{R}^L$ of dimension $r < L$. It is assumed that $e_L \notin \mathfrak{L}_r$, where $e_L = (0, 0, \ldots, 1) \in \mathbf{R}^L$.

(d) Number M of points to forecast.

Notations and comments:

(a) $\mathbf{X} = [X_1, \ldots, X_K]$ is the trajectory matrix of the time series Y_T.

(b) $P_1, \ldots, P_r$ is an orthonormal basis in $\mathfrak{L}_r$.

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

(d) $\widetilde{\mathbf{X}} = \mathcal{H}\mathbf{X} = [\widetilde{X}_1 : \ldots : \widetilde{X}_K]$ is the result of the Hankellization of the matrix $\widehat{\mathbf{X}}$.

(e) For any vector $Y \in \mathbf{R}^L$ we denote by $Y_\Delta \in \mathbf{R}^{L-1}$ the vector consisting of the last $L-1$ components of the vector Y, while $Y^\nabla \in \mathbf{R}^{L-1}$ is the vector of the first $L-1$ components of the vector Y.

(f) We set $v^2 = \pi_1^2 + \ldots + \pi_r^2$, where π_i is the last component of the vector P_i $(i = 1, \ldots, r)$.

(g) Suppose that $e_L \notin \mathfrak{L}_r$(This implies that $\mathfrak{L}_r$ is not a vertical space). Then $v^2 < 1$. It can be proved that the last component y_L of

any vector $Y = (y_1, \ldots, y_L)^T \in \mathfrak{L}_r$ is a linear combination of the first components $(y_1, \ldots, y_{L-1})$:

$$y_L = a_1 y_{L-1} + \ldots + a_{L-1} y_1.$$

Vector $A = (a_1, \ldots, a_{L-1})$ can be expressed as

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

and does not depend on the choice of a basis $P_1, \ldots, P_r$ in the linear space $\mathfrak{L}_r$. In the above notations, define the time series $Y_{T+M} = (y_1, \ldots, y_{T+M})$ by the formula

$$y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, T \\ \sum_{j=1}^{L-1} a_j y_{i-j} & \text{for } i = T+1, \ldots, T+M \end{cases} \tag{2.4.2}$$

The numbers $y_{T+1}, \ldots, y_{T+M}$ from the M terms of the SSA recurrent forecast. Let us define the linear operator $\mathcal{P}^{(r)} : \mathfrak{L}_r \mapsto \mathbf{R}^L$ by the formula

$$\mathcal{P}^{(r)} Y = \begin{pmatrix} Y_\Delta \\ A^T Y_\Delta \end{pmatrix}, \quad Y \in \mathfrak{L}_r$$

Set

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, K \\ \mathcal{P}^{(r)} Z_{i-1} & \text{for } i = K+1, \ldots, K+M \end{cases} \tag{2.4.3}$$

the matrix $\mathbf{Z} = [Z_1, \ldots, Z_{K+M}]$ is the trajectory matrix of the series Y_{T+M}. Therefore, (2.4.3) can be regard as the vector form of (2.4.2).

2.5 Bootstrapping

Assume that we have a time series $Y_T = \{y_t\}_{t=1}^T = Y_T^{(1)} + Y_T^{(2)}$, where $Y_T^{(1)}$ is the signal and $Y_T^{(2)}$ represents the noise. Let us consider a method of constructing average series for the signal $y_{T+M}^{(1)}$ at time $T+M$. In the unrealistic situation, when we know both the signal $Y_T^{(1)}$ and the true model of the noise $Y_T^{(2)}$, the Monte Carlo simulation can be applied to check the statistical properties of the forecast values $\widetilde{y}_{T+M}^{(1)}$ relative to the actual term $y_{T+M}^{(1)}$.

Indeed, assuming that the rules for the eigentriple selection are fixed, we can simulate N independent copies $Y_{T,i}^{(2)}$ $(i = 1, \ldots, N)$ of the process $Y_T^{(2)}$ and apply the forecasting procedure to N independent time series $Y_{T,i} = Y_T^{(1)} + Y_{T,i}^{(2)}$. Then the forecasting result will form a sample $\widetilde{y}_{T+M,i}^{(1)}$, which should be compared against $y_{T+M}^{(1)}$. In this way the Monte Carlo average series for the forecast can be built up.

Since in practice we do not know the signal $Y_T^{(1)}$, we can not apply this procedure. Under a suitable choice of the window length L and the corresponding eigentriples, we have the representation $Y_T = \widetilde{Y}_T^{(1)} + \widetilde{Y}_T^{(2)}$, where $\widetilde{Y}_T^{(1)}$ (the reconstructed series) approximates $Y_T^{(1)}$, and $\widetilde{Y}_T^{(2)}$ is the residual series. Suppose now that we have a (stochastic) model for the residual $\widetilde{Y}_T^{(2)}$ (for instance, we can postulate some model for $Y_T^{(2)}$ and, since $\widetilde{Y}_T^{(1)} \approx Y_T^{(1)}$, we apply the same model for $\widetilde{Y}_T^{(2)}$ with the estimated parameters). Then, simulating N independent copies $Y_{T,i}^{(2)}$ of the series $\widetilde{Y}_T^{(2)}$, we obtain N series $Y_{T,i} = \widetilde{Y}_T^{(1)} + \widetilde{Y}_{T,i}^{(2)}$ and produce M forecasting results $\widetilde{y}_{T+M,i}^{(1)}$ in the same manner as in the Monte Carlo simulation variant.

From the sample $\widetilde{y}_{T+M,i}^{(1)} (1 \leq i \leq N)$ of the forecasts we can compute the average bootstrap forecast. This average bootstrap can then be

compared with the value $\widetilde{y}_{T+M}^{(1)}$ obtained by Basic SSA forecast. Large discrepancy between these two forecasts would typically indicate that the original SSA forecast is not reliable. Furthermore, using the sample of the bootstrap forecast results we can estimate the distribution of the forecasts and compute, for example, confidence intervals for the true values. To do that, we need a stochastic model for $Y_T^{(2)}$; a standard assumption would be the assumption that $Y_T^{(2)}$ is a Gaussian white noise. This assumption can be easily verified using the classical tests for randomness and normality.

2.6 Confidence intervals for the forecasts

Confidence intervals for the forecasts can be calculated by two methods: the empirical method and the bootstrap method (which is also an empirical method). They are calculated using the residuals of the reconstruction.

According to the main SSA forecasting assumption, the component $Y_T^{(1)}$ of the series Y_T has to satisfy an LRF of a relatively small dimension, and the residual series $Y_T^{(2)} = Y_T - Y_T^{(1)}$ has to be approximately separable from $Y_T^{(1)}$. In particular, $Y_t^{(1)}$ is assumed to be a finite subseries of an infinite series $Y^{(1)}$, which is a recurrent continuation of $Y_T^{(1)}$. These assumptions are often hold in practice with high accuracy.

There are two problems related to the construction of the confidence intervals for the forecast. The first problem is to construct a confidence interval for the original series $Y_T = \{y_t\}$ at some future point in time. The second problem is construction of confidence intervals for the signal $Y_T^{(1)} = \{y_t^{(1)}\}$ at some future point in time. These two problems can be solved in different ways. The second requires additional infor-

mation about the model governing the series $\widetilde{Y}_T^{(2)} = \{\widetilde{y}_t^{(2)}\}$ to perform a bootstrap simulation of the series Y_T. Bootstrap confidence intervals are built for the continuation of the signal $Y_T^{(1)}$ (for more information see [8]).

Let us consider a method of constructing intervals for the signal $Y_{T+M}^{(1)}$ at the moment $T+M$. In the unrealistic situation, when we know both the signal $Y_T^{(1)}$ and the true model of the noise $Y_T^{(2)}$, a Monte Carlo simulation can be applied to check the statistical properties of the forecast value $\widetilde{y}_{T+M}^{(1)}$ relative to the actual term $y_{T+M}^{(1)}$.

Indeed, assuming that the rules for the eigentriple selection are fixed, we can simulate N independent copies $Y_{T,i}^{(2)}$ $(i = 1, 2, \cdots, N)$ of the process $Y_T^{(2)}$ and apply the forecasting procedure to N independent time series $Y_{T,i} = Y_T^{(1)} + Y_{T,i}^{(2)}$. Then the forecasting result will form a sample $\widetilde{y}_{T+M,i}^{(1)}$, which should be compared against $y_{T+M}^{(1)}$. In this way the Monte Carlo average series for the forecast can be built up. Since in practice we do not know the signal $Y_T^{(1)}$, we can not apply this procedure. Let us describe the bootstrap variant of the simulation for constructing the confidence intervals for the forecast.

Under a suitable choice of the window length L and the corresponding eigentriples, we have the representation $Y_T = \widetilde{Y}_T^{(1)} + \widetilde{Y}_T^{(2)}$, where $\widetilde{Y}_T^{(1)}$ (the reconstructed series) approximates $Y_T^{(1)}$, and $\widetilde{Y}_T^{(2)}$ is the residual series. Suppose now that we have a (stochastic) model for the residual $\widetilde{Y}_T^{(2)}$ (for instance, we can postulate some model for $Y_T^{(2)}$ and, since $\widetilde{Y}_T^{(1)} \approx Y_T^{(1)}$, we apply the same model for $\widetilde{Y}_T^{(2)}$ with the estimated parameters). Then, simulating N independent copies $Y_{T,i}^{(2)}$ of the series $\widetilde{Y}_T^{(2)}$, we obtain N series $Y_{T,i} = \widetilde{Y}_T^{(1)} + \widetilde{Y}_{T,i}^{(2)}$ and produce M forecasting results $\widetilde{y}_{T+M,i}^{(1)}$ in the same manner as in the Monte Carlo simulation

variant.

More precisely, any time series $Y_{T,i}$ produces its own $\widetilde{Y}_{T,i}^{(1)}$ reconstructed series and its own forecasting linear recurrent formula LRF_i for the same window length L and the same sets of eigentriples. Starting at the last $L-1$ terms of the series $\widetilde{Y}_{T,i}^{(1)}$, we perform M steps of forecasting with the help of its LRF_i, to obtain $\widetilde{y}_{T+M,i}^{(1)}$.

From the sample $\widetilde{y}_{T+M,i}^{(1)} (1 \leq i \leq N)$ we can calculate its (empirical) lower and upper quintiles for a fixed level γ and obtain the corresponding confidence interval for the forecast. This interval (called bootstrap confidence interval) can be compared with the forecast value $\widetilde{y}_{T+M}^{(1)}$ obtained from the initial forecasting procedure. We can also build average bootstrap series. This average can then be compared with the value $\widetilde{y}_{T+M}^{(1)}$ obtained by Basic SSA forecast. Large discrepancy between these two forecast would typically indicate that the original SSA forecast is not reliable.

The simplest model for $\widetilde{Y}_T^{(2)}$ is the Gaussian white noise model. The corresponding hypothesis can be checked with the help of the standard test for randomness and normality.

2.7 Multivariate singular spectrum analysis (MSSA)

The use of multivariate singular spectrum analysis (MSSA) for multivariate time series was proposed theoretically in the context of nonlinear dynamics in [9]. There are numerous examples of successful application of the multivariate SSA (see, for example, [1] and [10]). Multivariate (or multichannel) SSA is an extension of the standard SSA to the case of multivariate time series. We give a short description of MSSA method as follows.

Assume that we have an M-variate time series $y_j = (y_j^{(1)}, \ldots, y_j^{(M)})$, where $j = 1, \ldots, T$ and let L be window length.

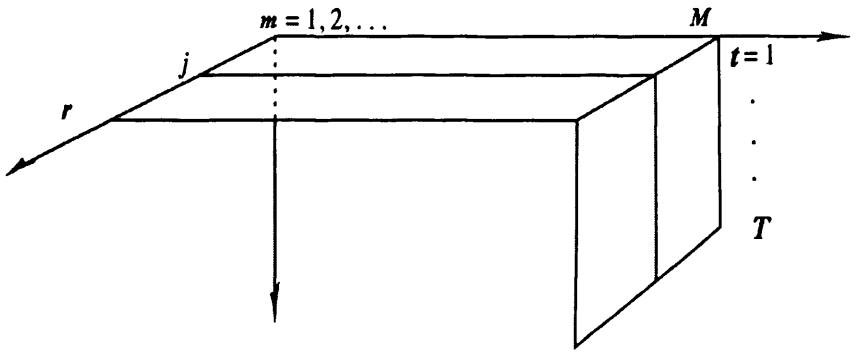


Figure 2.1. An illustration of MSSA.

Similar to univariate version, we can define the trajectory matrices $\mathbf{X}^{(i)}$ $(i = 1, \ldots, M)$ of the one-dimensional time series $\{y_j^{(i)}\}$ $(i = 1, \ldots, M)$. The trajectory matrix $\mathbf{X}$ can then be defined as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(M)} \end{pmatrix}. \tag{2.7.1}$$

Fig. 2.1 shows an illustration of MSSA. The structure of matrix $\mathbf{C} = \mathbf{X}\mathbf{X}^T$ is as follows:

$$\mathbf{C} = \begin{pmatrix} C_{11} & \ldots & C_{1m} & \ldots & C_{1M} \\ \vdots & \ddots & \vdots & \ldots & \vdots \\ C_{m1} & \ldots & C_{mm} & \ldots & C_{mM} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{M1} & \ldots & C_{Mm} & \ldots & C_{MM} \end{pmatrix}, \tag{2.7.2}$$

where, $C_{IJ} = \mathbf{X}^{(I)}(\mathbf{X}^{(J)})^T$ $(I, J = 1, \ldots, M)$ is an estimate of the covariance between two trajectories $\mathbf{X}^{(I)}$ and $\mathbf{X}^{(J)}$ corresponding to the

series Y^I and Y^J. The other stages of multivariate SSA procedure are identical to the basic SSA as described above with an obvious modification that the diagonal averaging should be applied to each of the M components separately.

Chapter 3

SSA AS A NOISE REDUCTION METHOD

In this chapter, the daily closing prices of several stock market indices are examined to analyse whether noise reduction matters in measuring dependencies of the financial series. We consider the effect of noise reduction on the degree of the linear and nonlinear measure of dependencies between to time series. We also use SSA as a powerful method for filtering financial series. The results are compared with those obtained by ARMA and GARCH models as linear and nonlinear methods for filtering the series. We also examine the findings on an artificial data set namely the Hénon map.

3.1 Introduction

During the last few years the analysis of financial time series has received increasing attention. Many researchers have discovered evidence for the possibility that the financial markets may be nonlinear dynamical systems, with important implications in the Efficient Market Hypothesis. Several researchers, by using different statistical tests, have mentioned evidence of non-independently and identically distributed

behaviour of the financial time series, and also the existence of the nonlinear dependence among these series [35]– [43].

Several measures have been used to calculate the degree of independency or dependency. The most known measure to calculate dependency between two random variables is the coefficient of linear correlation, but its application requires a pure linear relationship, or at least a linear transformed relationship. This statistics may not be helpful in determining serial dependence if there is some kind of nonlinearity in the data [44, 45].

Urbach [46] defends a strong relationship between entropy, dependence and predictability. This relation has been studied by many authors [45]– [48]. It has been shown that a measure based on the mutual information, which captures linear and nonlinear dependencies, without requiring the specification of any kind of model of dependence, is better than the linear correlation coefficient to measure serial correlation of several stock market indices [44]– [48].

Recently, two new methods have been developed to measure long-range correlations in non-stationary fluctuating series; the detrended fluctuation analysis [49, 50] and the detrended moving average method [51,52]. These methods detect persistency by assuming the self-similarity of the series.

It is well known that the existence of a significant noise level reduces the efficiency of the methods to analyze financial time series. Consider a time series $y_t = s_t + \epsilon_t$ $(t = 1, \ldots, T)$ which behaves as stochastic dynamic systems with both a deterministic element, s_t, and a stochastic part ϵ_t. We consider the second part as noise. Here we investigate the efficiency of noise reduction on the measures of dependencies (linear

and nonlinear).

We mainly follows two different approaches to calculate the measures of dependence. According to the first one, we calculate the measures of dependencies directly from the noisy time series. Therefore, we ignore the existence of the noise in the first approach. According to the second approach we start with filtering the noisy time series in order to reduce the noise level and then calculate the measures. It is clear that the results by the second approach are more effective than the first one if we select a proper method for filtering the series.

There are several nonlinear methods for filtering noisy series such as local projective, Digital Butterworth filters, splines, filters based on spectral analysis, singular value decomposition (SVD) and simple nonlinear filtering. It has been shown that the SVD-based methods are more effective than the other ones for the reduction of noise in financial time series [53]. Here, we use the SSA technique as a tool for filtering financial time series. Recent research shows that SSA can be used as an alternative to traditional filtering methods [31]. For example, Alonsoa [32] showed superiority of the SSA technique over traditional methods used in biomechanical analysis for filtering data. Moreover, it has been shown that SSA can be used as a filtering method for longitudinal data and growth curve models [34]. Here we will show that one should consider at least two criteria to capture the values of dependencies for both a single series and for a set of noisy financial series; one is selecting a proper method for filtering data (e.g. the SSA technique), and the other is considering a measure which can capture the considerable values of nonlinearity of the series.

3.2 Linear and nonlinear dependency

If one considers financial series to be deterministic as linear dynamic systems, a linear measure of dependencies such as linear correlation can be used for measuring dependencies between two time series. Most financial models are based on the assumption of multivariate normality (for example modelling dependent risks) and linear correlation is used as a measure of dependence. However, observed financial data are rarely normally distributed and tend to have marginal distributions with heavier tails [54].

The absence of economically significant linear correlations in price increments and asset returns has been widely documented (see [55] and references therein) and often cited as support for the Efficient Market Hypothesis [56]. It is also a well-known fact that price movements in liquid markets do not exhibit significant autocorrelation.

Alternatively, it has been shown that most of the financial time series are nonlinear (see, for example, [53, 57, 58]). Based on this scenario, we should use measures which have the capability to capture the nonlinearities of the series. Granger and Lin [45], and Darbellay and Wuertz [47] defined a standard measure based on the mutual information which can be used to capture the nonlinearities in the financial time series.

Moreover, nonstationarity can often be associated with different trends in the signal or heterogeneous segments with different local statistical properties. To address this problem, detrended fluctuation analysis (DFA) and detrended moving average (DMA) were developed to accurately quantify long-range power-law correlations embedded in a nonstationary time series [49]– [51].

3.3 Empirical Results

We shall consider two types of time series; real and artificially generated time series, financial data and a Hénon map, respectively. First we consider the values of autocorrelation function (ACF), nonlinear correcation based on the mutual information (λ[1]), the DFA α exponent (α_{DFA}), and the DMA α exponent (α_{DMA}) before and after noise reduction (for more details see Appendix C). To evaluate the performance of the SSA technique for filtering these series, we also use a linear and a nonlinear method, namely the ARMA and GARCH models which are used in [48] and [59] for filtering the same financial series used here. For more information about the ARMA and GARCH models see Appendix C.

3.3.1 Hénon map

The capability of the SSA technique as a noise reduction method for filtering chaotic time series was initially tested by applying the technique to the Hénon map [60]:

$$\begin{cases} x_{t+1} = 1 + y_t - A\,x_t^2 \\ y_{t+1} = B\,x_t \end{cases} \tag{3.3.1}$$

with usual parameter values: $A = 1.4$ and $B = 0.3$. In total 1895 data points are generated and we add different normally distributed noise to each point of the original series.

Table 3.2 represents the results of the ACF at lag 1, λ, α_{DFA}, and α_{DMA} before and after filtering. Y_{T_1} has the smallest noise level and

[1] $\lambda = \left(1 - exp[-2I(X,Y)]\right)^{\frac{1}{2}}$, where $I(X,Y)$ is the mutual information of two series X and Y. For more information see Appendix C.

Y_{T_4} the largest. The values of the ACF at lag 1, λ, α_{DFA}, and α_{DMA} of noise-free Hénon map are -0.335, 0.601, 0.373 and 0.585, respectively. The first row in each panel (labeled Noisy Hénon), shows the value of the ACF, λ, α_{DFA} and α_{DMA} before filtering. Here we used MATLAB to calculate these quantities. Let us first consider the value of w-correlation (see Appendix F) for different values of L and r. Table 3.1 shows the results. As the results indicate, different combinations of L and r yields different orthogonality results.

	L		
	10	100	400
$r = 0.1L$	0.201	0.066	0.103
$r = 0.5L$	0.212	0.093	0.113
$r = 0.9L$	0.387	0.027	0.061

Table 3.1. The value of w-correlation for different values of L and r.

	Y_{T_1}	Y_{T_2}	Y_{T_3}	Y_{T_4}
	ACF			
Noisy Hénon	-0.375	-0.378	-0.383	-0.388
ARMA	-0.023	-0.024	-0.022	-0.026
GARCH	-0.032	-0.033	-0.039	-0.045
SSA	-0.337	-0.339	-0.343	-0.345
	λ			
Noisy Hénon	0.970	0.937	0.905	0.857
ARMA	0.812	0.783	0.703	0.656
GARCH	0.856	0.857	0.856	0.720
SSA	0.634	0.667	0.640	0.617
	α_{DFA}			
Noisy Hénon	0.364	0.355	0.421	0.402
ARMA	0.415	0.410	0.471	0.430
GARCH	0.409	0.401	0.402	0.425
SSA	0.377	0.379	0.385	0.392
	α_{DMA}			
Noisy Hénon	0.578	0.580	0.625	0.638
ARMA	0.581	0.582	0.630	0.636
GARCH	0.581	0.582	0.589	0.634
SSA	0.582	0.583	0.587	0.592
Quantities	ACF	λ	α_{DFA}	α_{DMA}
Hénon	-0.335	0.601	0.373	0.585

Table 3.2. The values of the ACF at lag 1, λ, α_{DFA}, and α_{DMA} of the Hénon map for different noise levels.

As appears from Table 3.2, different noise levels give different values of the ACF, λ, α_{DFA}, and α_{DMA}. Table 3.2 shows that the estimated values of the ACF, λ, α_{DFA}, and α_{DMA}, after noise reduction, based on the SSA technique are more robust than the other methods that are considered here.

The results also indicate that the values of λ after filtering by ARMA, GARCH and SSA are more accurate than the values of the noisy series. Confirming the existing results in filtering financial data literature [61], the results in Table 3.2 show that nonlinear structure in chaotic series cannot be extracted properly with a GARCH model.

As can be seen from Table 3.2, the value of α_{DMA} increases as the noise level increases, while we do not observe this for α_{DFA}. This means that the DMA is more sensitive than the DFA regarding different noise levels. Note also that, the value of $\alpha_{DFA} = 0.373$ indicates antipersistence, while $\alpha_{DMA} = 0.585$ indicates a low level of positive correlations in noise-free Hénon map. These results coincide with those obtained in previous works (for example, Grech and Mazur [62] showed that good concurrence between the DFA and the DMA methods is found for long time series, $T \sim 10^5$, while for shorter series discordant results obtained for two methods with no systematic relation between them). It should be noted that the time series obtained from stochastic (noise-driven) and deterministic systems may be indistinguishable using the DFA method [63].

3.3.2 Financial series

From the data base DataStream, we selected the daily closing prices of several stock market indices: ASE (Greece) , CAC 40 (France), DAX 30

(Germany), FTSE 100 (UK), PSI 20 (Portugal), IBEX 35 (Spain) and S&P 500 (USA), spanning the period from 2/01/1990 to 28/09/2007 (which corresponds to 4629 observations per index), in order to compute the rates of return. These data sets have been used by several authors (see, for example, [48] and [59]). Szpiro [64], in studying the S&P 500 Index, found an increasing presence of noise. Davis and Mikosch [65] consider plots of the sample ACF of the squares of the S&P index for different periods and found that either the process is non-stationary or that the process exhibits heavy tails. Figure 3.1 shows the series.

The BDS test [35] for nonlinearity was used to test whether the series are IID. The results of the BDS test indicates significant dependence in all series confirming the existing results of dependencies in stock market literature [36], [59].

Table 3.3 represents a summary of descriptive statistics for the series before and after filtering. The rows related to Kurtosis shows the value of Kurtosis of the series. A positive value typically indicates that the distribution has a sharper peak, thinner shoulders, and fatter tails than the normal distribution. As it appears from Table 3.3, all series have fatter tails than the normal distribution. Thus, the GARCH model was considered as a noise reduction method for filtering the series. As can be observed from Table 3.3, the filtered series based on the SSA, for all cases have a smaller standard deviation, S.D, than those values obtained by the GARCH model confirming the results obtained for the Hénon map. The same results can also be seen for the values of the maximum and minimum of the series. Note that here we used GARCH(1,1), and the w-correlation is about 0.07.

As we mentioned above, the DFA and the DMA present several

Statistics	Method	DAX 30	CAC 40	FTSE 100	IBEX 35	S&P 500	PSI 20	ASE
Mean$\times 10^{-3}$	Original	0.24	0.28	0.24	0.36	0.35	0.21	0.58
	GARCH	-0.24	-0.21	-0.18	-0.30	-0.17	-0.21	0.84
	SSA	0.23	0.28	0.24	0.36	0.35	0.21	0.57
S.D$\times 10^{-1}$	Original	0.11	0.11	0.09	0.11	0.09	0.08	0.16
	GARCH	0.11	0.11	0.09	0.11	0.09	0.08	0.16
	SSA	0.09	0.09	0.07	0.09	0.09	0.07	0.13
Min$\times 10^{-1}$	Original	-0.92	-0.74	-0.53	-0.82	-0.70	-0.80	-0.96
	GARCH	-0.93	-0.74	-0.54	-0.82	-0.70	-0.74	-0.90
	SSA	-0.54	-0.49	-0.45	0.65	-0.60	-0.69	-0.80
Max$\times 10^{-1}$	Original	0.55	0.62	0.54	0.63	0.54	0.62	1.53
	GARCH	0.55	0.61	0.53	0.62	0.53	0.69	1.51
	SSA	0.47	0.61	0.40	0.48	0.53	0.42	1.26
Skewness	Original	-0.45	-0.20	-0.19	-0.32	-0.14	-0.46	0.26
	GARCH	-0.42	-0.18	-0.19	-0.28	-0.14	-0.29	0.29
	SSA	-0.42	-0.14	-0.22	-0.34	-0.13	-0.44	0.23
Kurtosis	Original	4.57	3.39	3.44	3.83	4.33	8.58	6.94
	GARCH	4.46	3.38	3.44	3.76	4.33	8.44	6.91
	SSA	3.81	3.44	3.64	3.77	4.30	6.93	6.10

Table 3.3. Descriptive statistics of several stock indices returns series before and after filtering.

	DAX 30	CAC 40	FTSE 100	IBEX 35	S&P 500	PSI 20	ASE
				ACF			
Original	0.0519*	0.0344*	0.0234*	0.0524*	0.0147	0.137*	0.147*
GARCH	0.0001	0.0000	0.0235	-0.0007	0.0145	-0.0001	-0.0001
SSA	0.1790*	0.1680*	0.1516*	0.2383*	0.0147	0.4406*	0.4505*
				λ			
Original	0.3079*	0.2358*	0.1508*	0.2564*	0.1540*	0.3502*	0.3157*
GARCH	0.2799*	0.1171*	0.1508*	0.5382*	0.1540*	0.7951*	0.2909*
SSA	0.2921*	0.2425*	0.2326*	0.2855*	0.1475*	0.4977*	0.5263*

Table 3.4. The values of the ACF at lag-1 and λ of several stock indices returns series before and after filtering.

potentialities in the analysis of nonstationary series. Since the financial series we considered here are stationary (we use rate of returns), we will not calculate α_{DFA} and α_{DMA} here. Table 3.4 shows the values of the ACF at lag-1 and λ of several stock indices returns series before and after filtering. As it appears from Table 3.4, the values of the ACF is changed after filtering. In fact, the values were immediately affected by filtering. Also it should be noted that the sign of the ACF of the series

IBEX 35, PSI 20 and ASE was changed from positive to negative after filtering by the GARCH model indicating that the performance of the GARCH model is not very good for filtering the series. It seems that the results obtained for λ after filtering, are more robust than those for the ACF.

We also used Ljung-Box Q-statistics to test whether the values obtained for the ACF, before and after filtering, are significantly different from zero; $*$ indicates significant results at the 1% level of significance. The results indicate that the values of the ACF of the original series and those obtained after filtering by the SSA (except for S&P) are statistically significant.

We also considered the significance test for λ. In order to perform the test we followed the method which has been introduced in [45]. The critical values have been simulated for the null distribution and found through simulation of critical values based on a white noise. The critical values for a number of sample sizes and different significant levels were presented in [59]. The symbol $*$ indicates the results at the 1% level of significance; the values of λ, before and after filtering, are statistically significant.

We examined the efficiency of noise reduction on the ACF and λ of a single variable so far. Next we consider the efficiency of noise reduction on the linear correlation, ρ^2, between two series, and also λ to find whether noise reduction matters when measuring dependencies, linear or non-linear, between two series. Table 3.5 represents the values of the ρ and λ before and after filtering. The results show that the

[2]It should be noted however, that a linear correlation might have a bad performance for heavier tailed data. The aim here is to examine whether noise reduction matters for this measure.

values of ρ were reduced after filtering for all cases. The values of ρ of the original series are greater than 90% for all cases confirming that ρ is not a reliable measure to capture dependence between two financial series. We also see significant discrepancies between the values of ρ for the original series and filtered series.

Table 3.5 also represents the results of λ between two financial series before and after filtering. It can be seen from Table 3.5 that the values of λ are more reliable than the values of ρ as expected. Again, the results show that λ is more robust than ρ under noise reduction as the results obtained either by SSA or GARCH model were not changed dramatically (whilst this happen for ρ). However, there is no significant discrepancy between two filtering methods. But, the results are quite different with those obtained from original data.

3.4 Conclusion

We considered the efficiency of noise reduction on the linear and non-linear measure of dependencies. We examined the efficiency of the noise reduction on the ACF, λ, α_{DFA} and α_{DMA} of a single series. The results show that ACF is not a suitable measure to capture dependencies of either financial or chaotic series while λ can be considered as a reliable measure (see Tables 3.2 and 3.4). We also observed that, the value of α_{DMA} increases as the noise level increases, while we do not observe this for α_{DFA}. This means that the DMA is more sensitive than the DFA regarding different noise levels.

We found that the proper selection of the filtering method matters to find the accurate values of the ACF, λ, α_{DFA} and α_{DMA} of a single series. The results with strong evidence show that SSA can be used as

		DAX 30	CAC 40	FTSE 100	IBEX 35	S&P 500	PSI 20
				ρ			
	Original	0.963					
CAC 40	GARCH	0.786					
	SSA	0.711					
	Original	0.965	0.948				
FTSE 100	GARCH	0.689	0.782				
	SSA	0.617	0.749				
	Original	0.921	0.958	0.942			
IBEX 35	GARCH	0.701	0.766	0.676			
	SSA	0.610	0.688	0.614			
	Original	0.948	0.960	0.978	0.950		
S&P 500	GARCH	0.449	0.427	0.413	0.386		
	SSA	0.421	0.414	0.399	0.387		
	Original	0.959	0.924	0.949	0.951	0.939	
PSI 20	GARCH	0.497	0.508	0.447	0.519	0.246	
	SSA	0.443	0.449	0.401	0.470	0.240	
	Original	0.937	0.935	0.921	0.915	0.936	0.906
ASE	GARCH	0.264	0.257	0.251	0.254	0.125	0.241
	SSA	0.240	0.244	0.231	0.242	0.139	0.248
				λ			
	Original	0.817					
CAC 40	GARCH	0.788					
	SSA	0.792					
	Original	0.702	0.798				
FTSE 100	GARCH	0.699	0.790				
	SSA	0.680	0.771				
	Original	0.754	0.800	0.697			
IBEX 35	GARCH	0.710	0.777	0.695			
	SSA	0.691	0.764	0.683			
	Original	0.521	0.471	0.469	0.451		
S&P 500	GARCH	0.491	0.457	0.471	0.425		
	SSA	0.477	0.455	0.452	0.418		
	Original	0.593	0.547	0.458	0.582	0.293	
PSI 20	GARCH	0.526	0.524	0.481	0.518	0.282	
	SSA	0.482	0.493	0.457	0.507	0.284	
	Original	0.319	0.355	0.296	0.411	0.171	0.284
ASE	GARCH	0.310	0.330	0.290	0.309	0.134	0.282
	SSA	0.317	0.306	0.291	0.281	0.167	0.274

Table 3.5. The values of ρ and λ of several stock market indices series.

a powerful noise reduction method for filtering either noisy financial or chaotic series (see Tables 3.2 and 3.4).

Note that dependence between time series is important in multivariate time series analysis. In economics, for example, elucidation of various causalities between time series is vital to forecasting and pre-

diction. Here, we also examined the efficiency of noise reduction on the measure of dependencies between two series. Again, the results show that noise reduction matters for linear measures of dependence, ρ. We found that λ gives the more reliable results than the ρ before and after filtering (see Table 3.5).

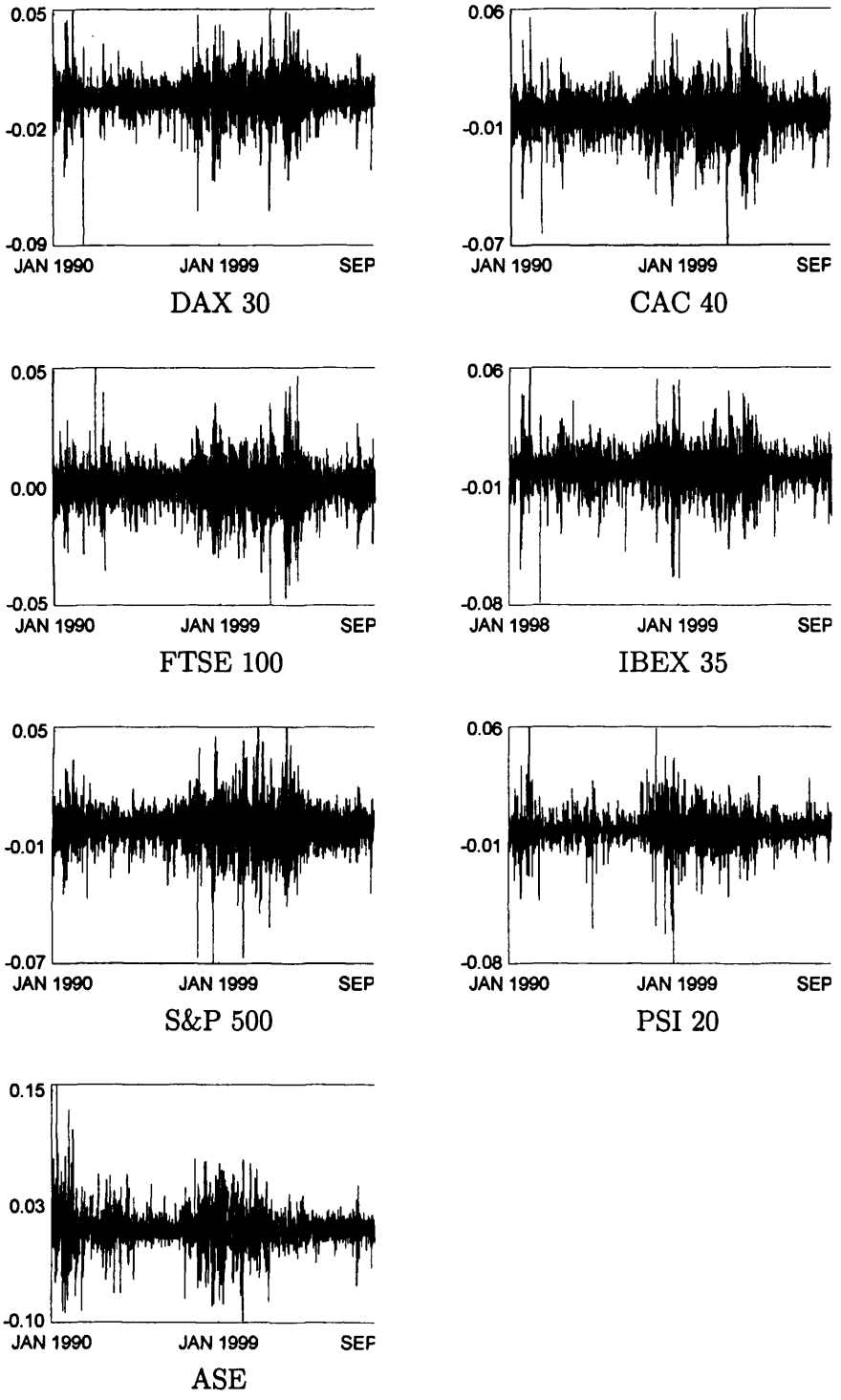


Figure 3.1. The daily closing prices of several stock indexes returns: DAX 30,CAC 40, FTSE 100, IBEX 35, S&P 500, PSI 20 and ASE.

Chapter 4

SSA AS A FORECASTING METHOD

In this chapter, the performance of the SSA technique in forecasting future data points is considered by applying it to several time series with various features, from short to long and from well structured to complex series. First, a well-known time series data set, namely, monthly accidental deaths in the USA, is used for detail analysis of the technique. A fully description of practical aspect of the method along with some criteria for selecting SSA parameters have been described by analyzing this series. The results of forecasting this series are compared with those obtained using Box-Jenkins SARIMA models, the ARAR algorithm and the Holt-Winter algorithm (as described in [3]).

Next, the performance of the SSA technique is assessed by applying it to 24 series measuring the monthly seasonally unadjusted industrial production for important sectors of the German, French and UK economies. The results are compared with those obtained using Holt-Winter and ARIMA models.

The application of SSA to the analysis and forecasting of short time series is evaluated using 32 Iranian national account data sets describing the main economic features of the Islamic Republic of Iran. The data

are given in a quarterly and yearly format and have different types of non-stationarity. All the data sets are rather short.

Moreover, the univariate and multivariate SSA (MSSA) are applied for predicting the value and the direction of changes in the daily pound/dollar exchange rate. The random walk model is used as a benchmark to evaluate performances of the SSA technique as a prediction method. The prediction results based on an error correction model in the context of a restricted vector autoregressive model are compared with the prediction results by a random walk as well as by those of SSA and MSSA.

The univariate and multivariate SSA are also employed in predicting inflation rate as well as the changes in direction of inflation time series for the United States. The consumer price indices, and real-time chain-weighted GDP price index series are used in these prediction exercises. Moreover, out-of-sample h-step-ahead moving prediction results are compared with the prediction results based on methods such as activity-based NAIRU Philips curve, $AR(p)$, and random walk models with the latter as a naive forecasting method. The short-run (quarterly) and long-run (one to six years) time windows are utilized for predictions. The results of earlier studies that indicates the prediction of inflation rate in the United States during the period of "Great Moderation" is less challenging compared to more volatile inflationary period of 1970-1985 is assessed using the results obtained by the SSA tecgnique.

4.1 American Death series

In the following we start with a brief description of the methodology of SSA and finish by applying it to the original series, namely, the monthly accidental deaths in the USA (Death series) and comparing the SSA technique with several other methods in forecasting this series.

4.1.1 The Data

The Death series shows the monthly accidental deaths in the USA between 1973 and 1978. This data have been used by many authors (see, for example, Brockwell and Davis [3]) and can be found in many time series data libraries. We apply the SSA technique to this data set to illustrate the capability of the SSA technique to extract trend, oscillation, noise and forecasting. All of the results and figures in the following application are obtained by means of the Caterpillar-SSA 3.30 software[1]. Fig. 4.1 shows the Death series over period 1973 to 1978.

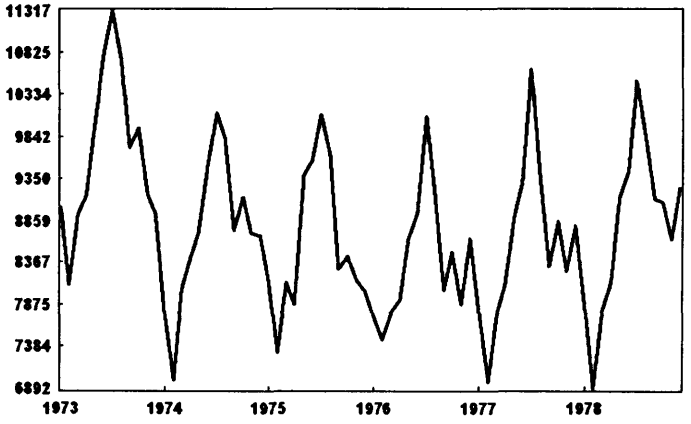


Figure 4.1. Death series: monthly accidental deaths in the USA (1973–1978).

[1]www.gistatgroup.com

Decomposition: Window Length and SVD

As we mentioned earlier, the window length L is the only parameter in the decomposition stage. Selection of the proper window length depends on the problem in hand and on preliminarily information about the time series. Theoretical results tell us that L should be large enough but not greater than $T/2$ [8]. Furthermore, if we know that the time series may have a periodic component with an integer period (for example, if this component is a seasonal component), then to get better separability of this periodic component it is advisable to take the window length proportional to that period. Using these recommendations, we take $L = 24$. So, based on this window length and on the SVD of the trajectory matrix (24×24), we have 24 eigentriples, ordered by their contribution (share) in the decomposition.

Note that the rows and columns of the trajectory matrix $\mathbf{X}$ are subseries of the original time series. Therefore, the left eigenvectors U_i and principal components V_i (right eigenvectors) also have a temporal structure and hence can also be regarded as time series (for further information see chapter 2). Let us consider the result of the SVD step. Fig. 4.2 represents the principal components related to the first 12 eigentriples. Note also that these principal components can be considered as a candidate in reconstruction stage. In fact, these components represent the structure of a subseries of the original series. For example in our case, the first principal component shows slowly varying pattern that can be considered as an evidence for reconstructing trend. The second and third principal components clearly show the harmonic pattern and therefore we consider these components for harmonic identification step.

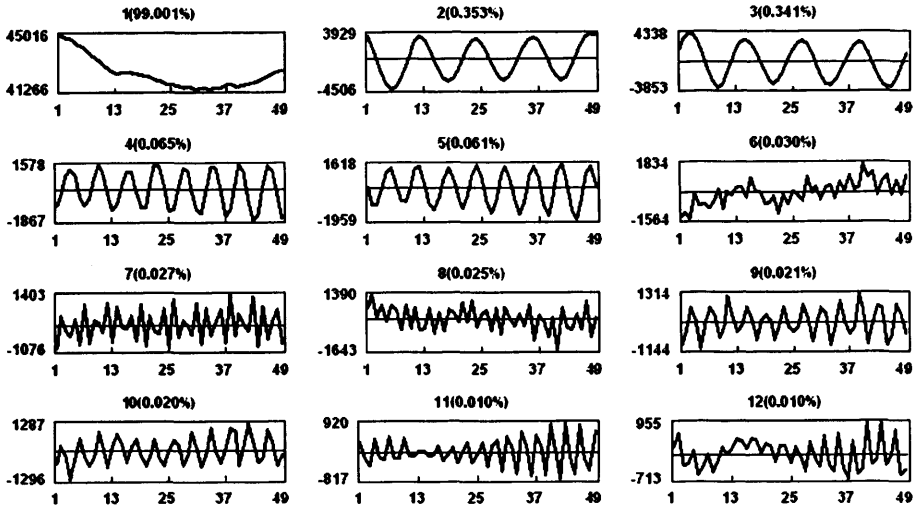


Figure 4.2. Principal components related to the first 12 eigentriples.

Supplementary Information

Let us describe some information, which proves to be very helpful in the identification of the eigentriples of the SVD of the trajectory matrix of the original series. Supplementary information help us to make the proper groups to extract the trend, harmonic components and noise. So, supplementary information can be considered as a bridge between the decomposition and reconstruction step:

$$\text{Decomposition} \longmapsto \text{Supplementary information} \longmapsto \text{Reconstruction}$$

Below, we briefly explain some methods, which are useful in the separation of the signal component from noise.

Auxiliary Information

The availability of auxiliary information in many practical situations increase the capability to build the proper model. Certainly, auxiliary information about the initial series always makes the situation clearer

and helps in choosing the parameters of the models. Not only can this information help us to select the proper group, but it is also useful for forecasting and the change point detection based on the SSA technique. For example, the assumption that there is an annual periodicity in the Death series suggests that we must pay attention to the frequency $k/12$ ($k = 1, ..., 12$). Obviously we can use the auxiliary information to select the proper window length as well.

Singular Values

Usually every harmonic component with a different frequency produces two eigentriples with close singular values (except for frequency 0.5 which provides one eigentriples with saw-tooth singular vector). It will be clearer if T, L and K are sufficiently large.

Another useful insight is provided by checking breaks in the eigenvalue spectra. As a rule, a pure noise series produces a slowly decreasing sequence of singular values.

Therefore, explicit plateaux in the eigenvalue spectra prompts the ordinal numbers of the paired eigentriples. Fig. 4.3 depicts the plot of the logarithms of the 24 singular values for the Death series.

Five evident pairs with almost equal leading singular values, correspond to five (almost) harmonic components of the Death series: eigentriple pairs 2-3 , 4-5, 7-8, 9-10 and 11-12 are related to harmonics with specific periods (we show later that they correspond to periods 12, 6, 2.5, 4 and 3).

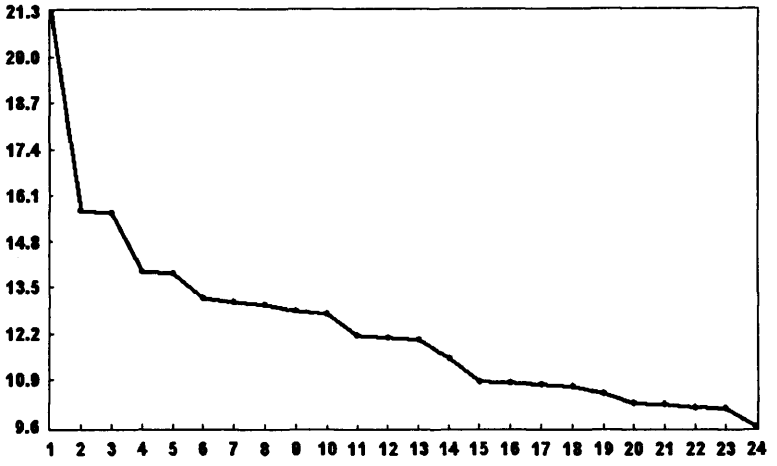


Figure 4.3. Logarithms of the 24 eigenvalues.

Pairwise Scatterplots

In practice, the singular values of the two eigentriples of a harmonic series are often very close to each other, and this fact simplifies the visual identification of the harmonic components. An analysis of the pairwise scatterplots of the singular vectors allows one to visually identify those eigentriples that corresponds to the harmonic components of the series, provided these components are separable from the residual component.

Consider a pure harmonic with a frequency w, certain phase, amplitude and ideal situation where $P = 1/w$ is a divisor of the window length L and K. Since P is an integer, it is a period of the harmonic. In the ideal situation, the left eigenvectors and principal components have the form of sine and cosine sequences with the same P and the same phase. Thus, the identification of the components that are generated by a harmonic is reduced to the determination of these pairs.

The pure sine and cosine with equal frequencies, amplitudes, and phases create the scatterplot with the points lying on a circle. If $P = 1/w$ is an integer, then this points are the vertices of the reg-

ular P-vertex polygon. For the rational frequency $w = m/n < 0.5$ with relatively prime integer m and n, the points are the vertices of the scatterplots of the regular n-vertex polygon. Fig. 4.4 depicts scatterplots of the 6 pairs of sine/cose sequence (without noise) with zero phase, the same amplitude and periods 12, 6, 4, 3, 2.5 and 2.4.

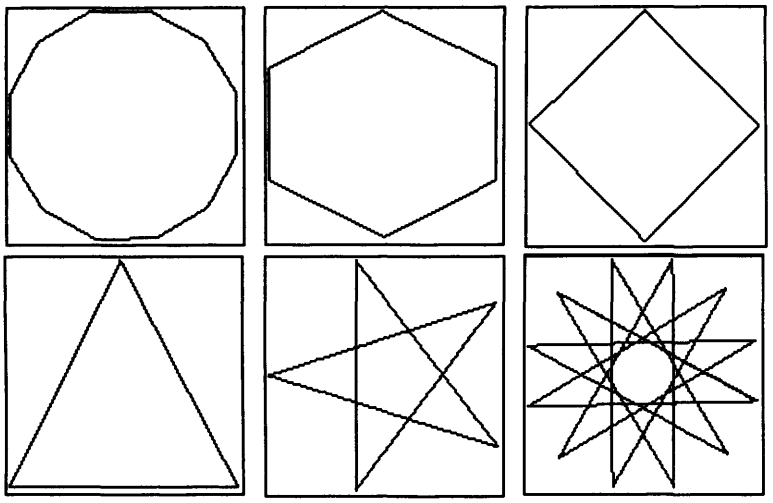


Figure 4.4. Scatterplots of the 6 pairs of sines/cosines.

Fig. 4.5 depicts scatterplots of the paired eigenvectors in the Death series, corresponding to the harmonics with periods 12, 6, 4, 3 and 2.5. They are ordered by their contribution (share) in the SVD step.

Periodogram Analysis

The periodogram analysis of the original series and eigenvectors may help us a lot in making the proper grouping; it tells us which frequency must be considered. We must then look for the eigentriples whose frequencies coincide with the frequencies of the original series.

If the periodograms of the eigenvector have sharp spark around some frequencies, then the corresponding eigentriples must be regarded as those related to the signal component.

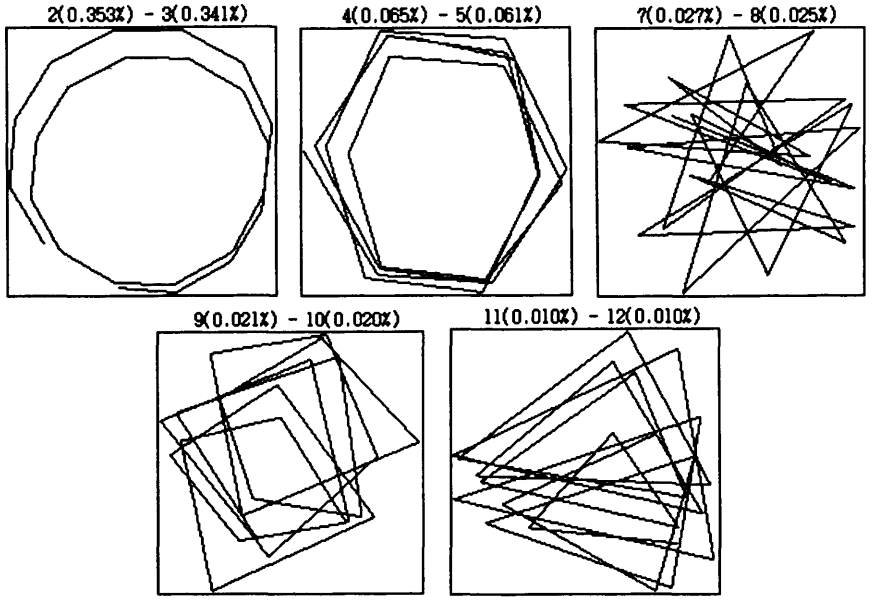


Figure 4.5. Scatterplots of the paired harmonic eigenvectors.

Fig. 4.6 depicts the periodogram of the paired eigentriples (2–3, 4–5, 7–8, 9–10, 11–12). The information arising from Fig. 4.6 confirms that the above mentioned eigentriples correspond to the periods $12, 6, 2.5, 4$ and 3 which must be regarded as selected eigentriples in the grouping step with another eigentriple we need to reconstruct the series.

Separability

The main concept in studying SSA properties is 'separability', which characterizes how well different components can be separated from each other. SSA decomposition of the series Y_T can only be successful if the resulting additive components of the series are approximately separable from each other. The following quantity (called the weighted correlation or *w-correlation*) is a natural measure of dependence between two series

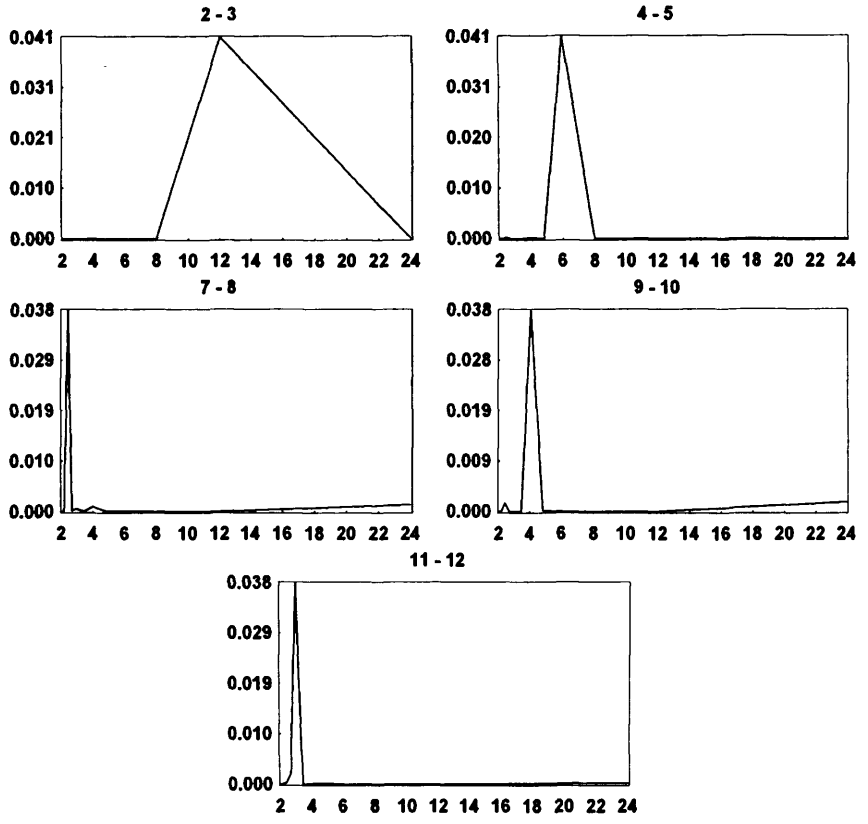


Figure 4.6. Periodograms of the paired eigentriples (2–3, 4–5, 7–8, 9–10, 11–12).

$Y_T^{(1)}$ and $Y_T^{(2)}$ [8]:

$$\rho_{12}^{(w)} = \frac{\left(Y_T^{(1)}, Y_T^{(2)}\right)_w}{\| Y_T^{(1)} \|_w \| Y_T^{(2)} \|_w}$$

where $\| Y_T^{(i)} \|_w = \sqrt{\left(Y_T^{(i)}, Y_T^{(i)}\right)_w}$, $\left(Y_T^{(i)}, Y_T^{(j)}\right)_w = \sum_{k=1}^{T} w_k y_k^{(i)} y_k^{(j)}$, $(i, j = 1, 2)$

w_k=min$\{k, L, T-k\}$ (here we assume $L \leq T/2$).

A natural hint for grouping is the matrix of the absolute values of the w-correlations, corresponding to the full decomposition (in this decomposition each group corresponds to only one matrix component of the SVD). If the absolute value of the w-correlations is small, then the corresponding series are almost w-orthogonal, but, if it is large, then

the two series are far from being w-orthogonal and are therefore weakly separable. So, if two reconstructed components have zero w-correlation it means that these two components are separable.

Fig. 4.7 shows the w-correlations for the 24 reconstructed components in a 20-grade grey scale from white to black corresponding to the absolute values of correlations from 0 to 1. Large values of w-correlations between reconstructed components indicate that the components should possibly be gathered into one group and correspond to the same component in SSA decomposition. In our case, there is almost two orthogonal blocks (eigenvalues 1–13 and eigenvalues 14–24). The w-correlation between these blocks is 0.004, indicating strong separability. Therefore, we can consider the reconstructed series obtained by eigenvalues 1–13 as signal and the rest as noise component.

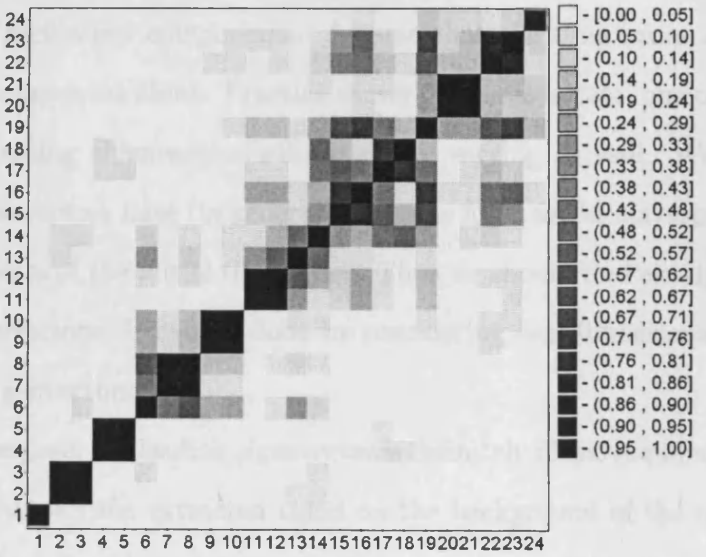


Figure 4.7. Matrix of w-correlations for the 24 reconstructed components.

Reconstruction: Grouping and Diagonal Averaging

Reconstruction is the second stage of the SSA technique. As mentioned above, this stage includes two separate steps: grouping (identifying signal component and noise) and diagonal averaging (using grouped eigentriples to reconstruct the new series without noise). Usually, the leading eigentriple describes the general tendency of the series. Since in most cases the eigentriples with small shares are related to the noise component of the series, we need to identify the set of leading eigentriples.

Grouping: Trend, Harmonics and Noise

Trend identification:

Trend is the slowly varying component of a time series which does not contain oscillatory components. Assume that the time series itself is such a component alone. Practice shows that in this case, one or more of the leading eigenvectors will be slowly varying as well. We know that eigenvectors have (in general) the same form as the corresponding components of the initial time series. Thus we should find slowly varying eigenvectors. It can be done by considering one-dimensional plots of the eigenvectors.

In our case, the leading eigenvector is definitely of the required form. Fig. 4.8 shows the extracted trend on the background of the original series which is obtained from the first eigentriple. Note that we can build a more complicated approximation of the trend if we use some other eigentriples. However, the precision we would gain will be very small and the model of the trend will become much more complicated.

Fig. 4.9 shows the extracted trend which is obtained from the first

and sixth eigentriples. It appears that taking the first and sixth eigentriples show the general tendency of the Death series better than the first eigentriple alone. However, the sixth eigentriple does not completely belong to the trend component but we can consider it as a mixture of the trend and the harmonic component.

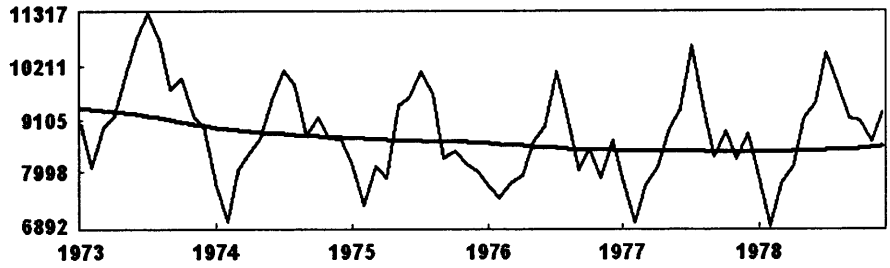


Figure 4.8. Trend extraction (first eigentriple).

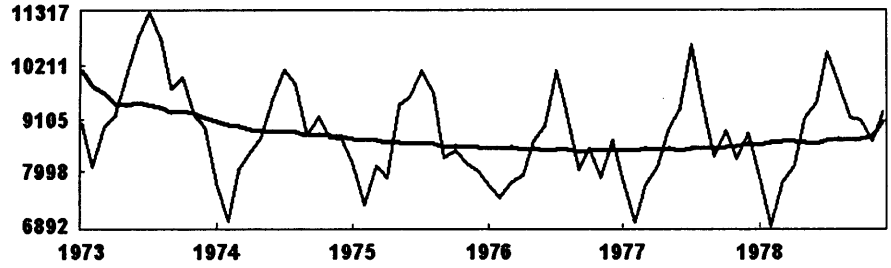


Figure 4.9. Trend extraction (first and sixth eigentriples).

Harmonic identification:

The general problem here is the identification and separation of the oscillatory components of the series that do not constitute parts of the trend. The statement of the problem in SSA is specified mostly by the model-free nature of the method.

The choice $L = 24$ allows us to simultaneously extract all the seasonal components (12, 6, 4, 3, and 2.5 month) as well as the trend. Fig. 4.10 shows the oscillation of our series which is obtained by the eigentriples 2-12.

By comparing Fig. 4.10 to Fig. 4.1 it is clear that the eigentriples selected to identify the harmonic components have been done so correctly. Fig. 4.11 shows the oscillation of our series obtained by the eigentriples 2–5 and 7–12. In this case we consider the sixth eigentriple as a trend component. It seems that there is no big discrepancy between selecting the sixth eigentriple into the trend or oscillation components as it appears from the Fig. 4.10 and 4.11.

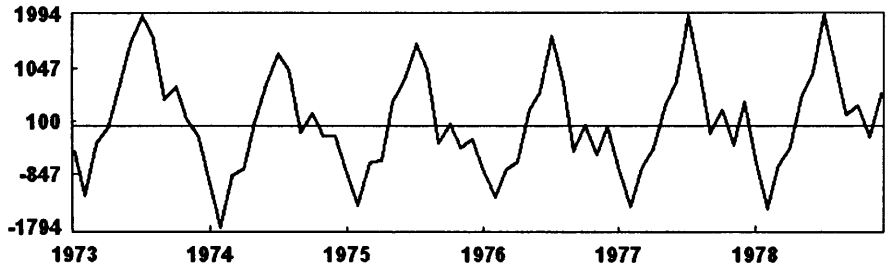


Figure 4.10. Oscillation extraction (eigentriples 2–12).

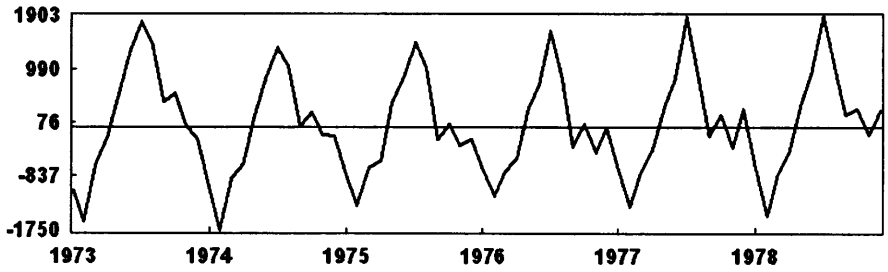


Figure 4.11. Oscillation extraction (eigentriples 2–5,7–12).

Noise detection:

The problem of finding a refined structure of a series by SSA is equivalent to the identification of the eigentriples of the SVD of the trajectory matrix of this series, which correspond to trend, various oscillatory components, and noise. From the practical point of view, a natural way of noise extraction is the grouping of the eigentriples, which do not seemingly contain elements of trend and oscillations. Let us discuss the eigentriple 13. We consider it as an eigentriple which belongs to noise because the period of the component reconstructed by eigentriple 13 is a mixture of the periods 3, 10, 14 and 24, as the periodogram indicates this cannot be interpreted in the context of seasonality for this series. We will thus classify eigentriple 13 as a part of the noise. Fig. 4.12 shows the residuals which are obtained by the eigentriples 13–24.

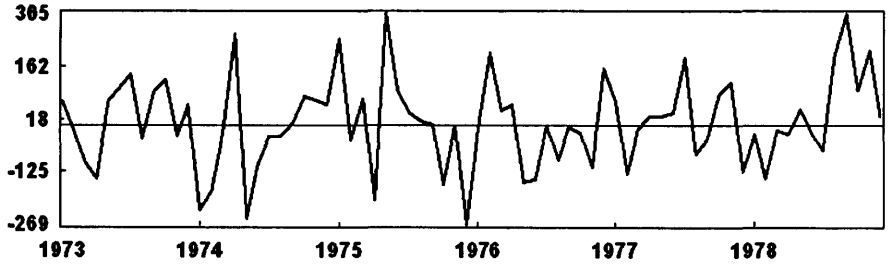


Figure 4.12. Residual series (eigentriples 13–24).

Diagonal Averaging

The last step of the SSA technique is diagonal averaging. If we just consider the trend (eigentriple 1 or (1 and 6)), harmonic component (eigentriple 2–12 or (2–5, 7–12)) and noise (eigentriple 13–24) as groups then we have 3 groups ($m = 3$). However we can have 8 groups if we consider each group by detail such as; eigentriples 1, 2–3, 4–5, 6, 7–8,

9-10, 11–12 (which correspond to the signal) and 13–24 or 7 groups if we merge the eigentriples 1 and 6 into a group. Fig. 4.13 shows the result of the signal extraction or reconstruction series without noise which is obtained from the eigentriples 1–12. The dotted and the solid line correspond to the reconstructed series and the original series respectively. As indicated on this figure, the considered groups for the reconstruction of the original series is optimal (bear in mind that the SVD step has optimal properties). If we add the series of Fig. 4.8 and 4.10 (or 4.9 and 4.11) we will obtain the refined series (Fig. 4.13).

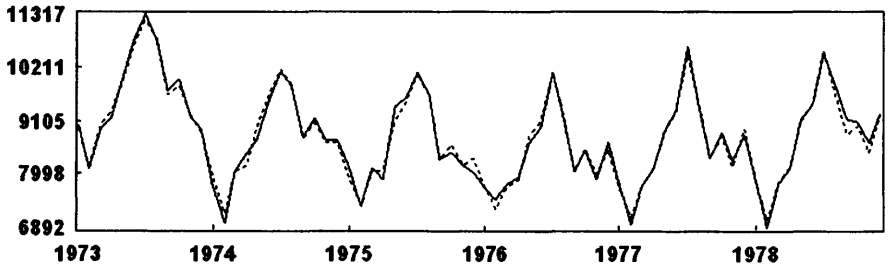


Figure 4.13. Reconstructed series (eigentriples 1–12).

Forecasting

Fig. 4.14 shows the original series (solid line), reconstructed series (dotted line) and its forecasting after 1978 (the six data points of 1979). The vertical dotted line shows the truncation between the last point of the original series and the forecast starting point. Fig. 4.14 shows that the reconstructed series (which is obtained from eigentriples 1-12) and the original series are close together indicating that the forecasted values are reasonably accurate.

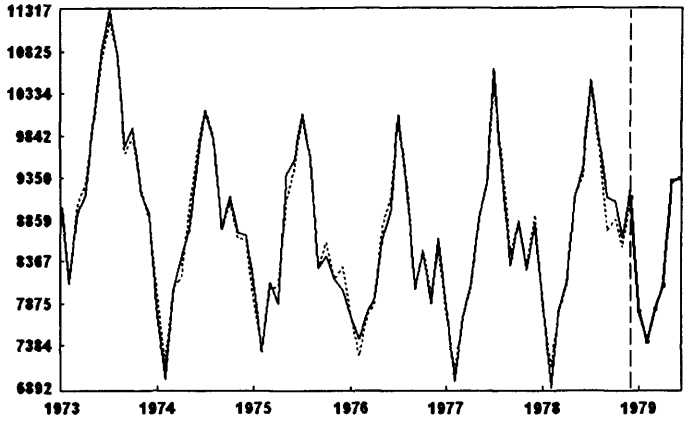


Figure 4.14. Original series (solid line), reconstructed series (dotted line) and the 6 forecasted data points of 1979.

4.1.2 Comparison

In this section we compare the SSA technique with several well-known methods namely, the traditional Box-Jenkins SARIMA models, the ARAR Algorithm and the Seasonal Holt-Winters Algorithm. Brockwell and Davis [3] applied these methods on the Death series to forecast the six future data points. Below, these methods are described shortly and the results of their forecasting are compared with the SSA technique.

SARIMA Model

Box and Jenkins [2] provide a methodology for fitting a model to an empirical series. This systematic approach identifies a class of models appropriate for the empirical data sequence at hand and estimates its parameters. A general class of Box and Jenkins models includes autoregressive moving average (ARIMA) and seasonal ARMA (SARIMA) models that can model a large class of autocorrelation functions. We use the models below for forecasting the six future data as are described in [3]:

Model I:

$$\nabla_{12} y_t = 28.831 + (1 - 0.478B)(1 - 0.588B^{12})Z_t, \quad Z_t \sim WN(0, 94390)$$

Model II:

$$\nabla_{12} y_t = 28.831 + Z_t - 0.596Z_{t-1} - 0.407Z_{t-6} - 0.685Z_{t-12} + 0.460Z_{t-13}$$

$$Z_t \sim WN(0, 94390)$$

where backward shift operator B is: $B^j Z_t = Z_{t-j}$.

Note that the seasonal difference of a time series is the series of changes from one season to the next. For monthly accidental deaths in the USA, in which there are 12 periods in a season, the seasonal difference of the series at period t is $\nabla_{12} y_t = y_t - y_{t-12}$. In the forecasting the series, we see that the first difference of y_t is far from random (it is still strongly seasonal), and the seasonal difference is far from stationary (it resembles a random walk). Therefore, both kinds of differencing are needed to render the series stationary and to account for the gross pattern of seasonality. It should be noted that the first difference of the seasonal difference of a monthly time series at period t is equal to $\nabla\nabla_{12} y_t$. This is the amount by which the change from the previous period to the current period is different from the change that was observed exactly one year earlier. Thus, for example, the first difference of the seasonal difference in May 1978 is equal to the April-to-May change in 1978 minus the April-to-May change in 1977.

ARAR Algorithm

The ARAR algorithm is an adaption of ARARMA algorithm (Newton and Parzen, 1984) in which the idea is to apply automatically selected 'memory-shortening' transformations (if necessary) to the data and then to fit an ARMA model to the transformed series. The ARAR algorithm used here is a version of this in which the ARMA fitting step is replaced by the fitting of the subset AR model to the transformed data.

Holt-Winter Seasonal Algorithm (HWS)

The Holt-Winter (HW) algorithm uses a set of simple recursions that generalize the exponential smoothing recursions to generate forecasts of series containing a locally linear trend. The Holt-Winter seasonal algorithm (HWS) extends the HW algorithm to handle data in which there are both trend and seasonal variation of known period.

Results

Table 4.1 shows the results for several methods for the forecasting of the six future data points. To calculate the precision we have used two measures, namely, the Mean Absolute Error (MAE) and the Mean Relative Absolute Error (MRAE) (for more information see Appendix A).

This table shows that the forecasted values are very close to the original data for the SSA technique. We borrow the result of forecasting for the other methods from Brockwell and Davis [3]. The methods are arranged based on the performance of forecasting. The values MAE and MRAE show the performance of forecasting (the value of the MRAE

	1	2	3	4	5	6	MAE	MRAE
Original Data	7798	7406	8363	8460	9217	9316		
Model I	8441	7704	8549	8885	9843	10279	524	6 %
Model II	8345	7619	8356	8742	9795	10179	415	5 %
HWS	8039	7077	7750	7941	8824	9329	351	4 %
ARAR	8168	7196	7982	8284	9144	9465	227	3 %
SSA	7782	7428	7804	8081	9302	9333	180	2 %

Table 4.1. Forecast data, MAE and MRAE for six forecasted data by several methods.

is rounded). As it appears in Table 4.1, the SSA technique is the best among the methods considered, for example, the value of MAE or MRAE for the SSA methods is 3 times less than the first one (model I) and 2 times less than the HWS algorithm.

Note that by using the above mentioned information and the SSA-Caterpillar software, anyone can repeat the results presented in this paper for each part such as the results of the forecasting in Table 4.1.

4.2 European Industrial Production

The SSA is especially useful for analyzing and forecasting series with complex seasonal components and non-stationarity. Thus, unlike ARIMA models, choosing an appropriate degree of differencing is not an important issue in SSA. The data considered in this study has a complex structure of this kind; as a consequence, we found that SSA is superior over classical techniques.

Here we use eight series of monthly industrial production indices for Germany, France and the UK, previously analysed in linear and nonlinear contexts in [66, 67]. The eight series examined for the three countries, Germany, France and the UK, are interesting and important since they cover production in the major industrial sectors. They also reflect diverse types of industries (see Table 4.2). Note that economic literature shows that it is possible to perform a model assessment tests on a small-area regional econometrics model, even though several highly informative tests are not commonly reported. In this case Theil-type U-statistics is useful [68].

Osborn et al. [66] have considered the extent and nature of seasonality in these series. Their findings show that seasonality accounts for over 90% of the variation in almost all French series. The strong seasonal pattern for the traditional industrial sector in France is associated with declines in production during the summer. Seasonality also accounts for at least 80% of variation in all series in Germany and in all series (except vehicles) in the UK. Osborn et al. [66] demonstrated that seasonalities for these series are much larger than those reported for monthly output in the United States at the two-digit level. The difference in pattern of seasonality between the European countries and

the United States is associated to differences in traditions and institutions. Based on seasonal unit root tests, Osborn et al. [66] found that most of the series should be modelled using conventional first difference. However, annual difference specification often produced the most accurate out-of-sample forecasts.

Heravi et al. [67] found relatively little evidence of non-linearity in most series. Comparing linear and neural network forecasts, they found that linear models generally produce more accurate post-sample forecasts than neural network models at horizons of up to a year in terms of root mean square error.

Here we examine the out-of-sample forecast accuracy of the SSA technique and compare it with ARIMA models and the Holt-Winter method.

4.2.1 The data

The data in this study are taken from Eurostat, the official statistical agency of the European Community and represents eight major components of industrial production in Germany, France and the UK. The series used are seasonally unadjusted monthly indices for real output in Food Products, Chemicals, Basic Metals, Fabricated Metals, Machinery, Electrical Machinery, Vehicles and Electricity/Gas industries. Appendix E provides detailed information about the series. It should be noted that the series for Germany are the aggregated data following the reunification of the former East Germany and West Germany.

The same 24 series, ending in December 1995, have been previously examined in [66, 67]. As explained in these papers, these time series have been chosen primarily because of their importance to industrial

production across the three countries. These eight time series account for at least half of total industrial production in each country. Plots of these time series are included in [66] and broadly represent a period of growth in the 1980s and stagnation or recession during the early 1990s. Here we have updated the data and in all cases the sample period ends in July 2007. However, the starting dates are different which reflects the availability of consistent data from Eurostat. The data for Germany starts from January 1978, for France starts from January 1990 and for the UK starts from 1998.

In all cases, the final two and a half years (30 observations) of data are retained for out-of-sample forecast accuracy tests. For comparability and in line with the usual convention for economic time series, all time series are analysed in the logarithmic form and all subsequent results refer to the time series after this transformation. The descriptive statistics for these series are given in Table 4.2. For Germany, the vehicles series has the highest volatility, which is more than twice than the volatility of the other series. Similarly, the vehicles series has the highest volatility for France. The UK data, generally, are less volatile with gas and electricity series having highest volatilities.

Almost all of the industrial production series have complex structure with nonlinear trends and complex seasonality. As SSA is generally well-suited for non-stationary series with complex trend and periodicities, our hope was that SSA would perform well for analyzing and forecasting industrial production series (for an example, see Appendix D).

Series	Mean			S.D.			Weight		
	UK	GR	FR	UK	GR	FR	UK	GR	FR
Food products	4.64	4.42	4.58	0.067	0.195	0.129	10.2	7.6	9.0
Chemicals	4.65	4.41	4.52	0.087	0.192	0.176	8.5	8.6	8.9
Basic metals	4.54	4.58	4.51	0.107	0.098	0.175	3.8	4.5	4.3
Fabricated metal	4.61	4.39	4.50	0.064	0.201	0.194	5.8	7.2	9.8
Machinery	4.63	4.51	4.55	0.078	0.152	0.163	7.5	13.6	8.6
Electrical machinery	4.47	4.37	4.57	0.105	0.256	0.138	3.0	5.6	3.9
Vehicles	4.64	4.29	4.39	0.133	0.315	0.405	4.7	10.4	7.1
Electricity and gas	4.62	4.48	4.54	0.176	0.172	0.204	6.7	6.5	9.6

Table 4.2. Descriptive statistics of the series.

4.2.2 Forecasting Results

Comparison of the accuracy of the forecasts

We consider forecasting performance of the SSA, ARIMA and Holt-Winter techniques at different horizons h, of up to a year. We provide results for $h = 1, 3, 6$ and 12 (months). We use the data up to the end of 2004 as training sample (to perform SSA decomposition and to estimate parameters of ARIMA and Holt-Winter models). Thus, with two and a half years of the out-of-sample data, we have $N = 30, 28, 25$ and 19 out-of-sample forecast errors at the horizons $h = 1, 3, 6$ and 12, respectively.

Here, we use the RMSE and the percentage of forecasts that correctly predict the direction of change to measure the forecast accuracy.[2] Note that if RRMSE < 1, then the SSA outperforms the other methods (either ARIMA or Holt-Winter).

In computing Box-Jenkins ARIMA forecasts, we need to choose the lags, the degree of differencing and the degree of seasonality (p, d, q), $(P, D, Q)_s$, where $s = 12$. To do that we use the maximum order of

[2]We have also computed other measures based on the magnitude of forecast errors, such as relative root mean absolute errors. These measures yield qualitatively similar results to RMSE; we thus do not report them.

lags, set by the software, and apply the Bayesian Information Criterion (BIC). Holt-Winter forecasts are also obtained by minimizing the BIC. The SSA parameters, the window length L and the number of eigentriples r, are chosen based on the eigenvalue spectra and separability (see Appendix D). The parameters (L, r) of the SSA and the orders $(p, d, q), (P, D, Q)_s$ of the ARIMA models are given when the models are estimated using data up to the end of 2004. Appendix D gives details of the analysis for fabricated metal series for Germany.

Tables 4.3, 4.4 and 4.5 show the in-sample RMSE and RMSE ratios and the out-of-sample RMSE ratios for the UK, France and Germany. Some summary statistics (average RMSE, RRMSE of SSA models to the Holt-Winter and ARIMA models for each country and horizon) are also given at the bottom of each table. The summary statistics are the RMSE and the RRMSE averages and the scores. The score is the number of times when SSA model yields lower RMSE. SSA has produced lower RMSE for all the series for the in-sample results [3].

The averages and the scores for 1-step ahead show that SSA forecasts are comparable with the forecasts obtained by ARIMA and Holt-Winter models. However, the performance of the SSA, relative to ARIMA and Holt-Winter models, improves for forecasting at the horizons greater than one. The scores also confirm that the SSA forecasts outperform the forecasts produced by the ARIMA and Holt-Winter models, particularly at longer horizons. For all the series and three countries (24 cases), SSA outperforms the ARIMA $16, 18, 22$ and 23 times at $h = 1, 3, 6$ and 12 horizons respectively. It also outperforms the Holt-Winter models $16, 19, 23$ and 23 times at $h = 1, 3, 6$ and 12

[3] SSA gives the highest R^2, although all three methods fit the data well in-sample, with $R^2 > 81\%$.

horizons.

Table 4.6 summarizes the results of forecasts by ARIMA, Holt-Winter and SSA for all series. This table shows that the quality of 1-step ahead forecasts are similar for ARIMA and SSA; Holt-Winter forecasts being slightly worse. The quality of SSA forecasts at horizons $h = 3, 6$ and 12 is much better than the quality of ARIMA and Holt-Winter forecasts. As h increases, the quality of ARIMA and Holt-Winter forecasts becomes worse; the standard deviation of the ARIMA and Holt-Winter forecasts increases almost linearly with h. The situation is totally different for the SSA forecasts: the quality of SSA forecasts is almost independent of the value of h (at least, in the range of values of h considered in the paper). This evidence serves as a confirmation of the following facts:

(i) most of the series considered here have a structure which can described via a deterministic trend and seasonality (for an example, see Appendix D);

(ii) this structure is well recovered by the SSA;

(iii) in most cases, the structure of the series is relatively stable as it is well kept by the series for at least 12 months starting at any point.

Note that in the ideal situation, when we have a series which is a sum of a deterministic component (fully recovered by SSA) and a random noise, the error of SSA forecast will be exactly the same at any horizon. For more information, see Chap. 2 in [8].

Using the modified Diebold-Marino statistics (see appendix A), given in [69], we test for the statistical significance of the results of the fore-

casts. The symbol $*$ in the table indicates the results at the 10% level of significance or less. Comparing the SSA forecasts with the ARIMA, SSA outperforms the ARIMA significantly $2, 12, 9$ and 19 times at $h = 1, 3, 6$ and 12 horizons respectively at 10% significance level or less. SSA also outperforms the Holt-Winter significantly $6, 13, 16$ and 19 times at $h = 1, 3, 6$ and 12 horizons respectively at 10% significance level or less. Similar results have also been found when comparing the bootstrap forecasts, called in the table BSSA (to obtain bootstrap average series we have replicated the series 1000 times). In fact, the scores for all the horizons in Tables 4.3, 4.4 and 4.5 show that both the SSA and bootstrap SSA methods have outperformed the ARIMA and Holt-Winter models exactly the same number of times (160 times out of the total number of 192 cases).

We have also used the forecast encompassing test [70]. The symbol $+$ indicates the results at the 10% level of significance or less. The results also confirm the superiority of the SSA, with 54% of cases significantly better at the 10% level of significance or less.

Cumulative distribution functions (c.d.f.) of the absolute values of the out-of-sample errors (for all eight series and 3 countries) obtained by SSA, ARIMA and Holt-Winter forecasts are presented in Fig. 4.15. If the c.d.f. graph produced by one method is strictly above the graph of another c.d.f., then we can conclude that the errors obtained by the first method are stochastically smaller than the errors for the second method.

Suppose that we consider two distributions A and B, characterized respectively by c.d.fs F_A and F_B. Then distribution B dominates distribution A stochastically at first order if, for any argument

y, $F_A(y) \geq F_B(y)$. Higher orders of stochastic dominance can also be defined. To this end, we define repeated integrals of the c.d.f of each distribution. Formally, we define a sequence of functions by the recursive definition:

$$D^1(y) = F(y), \quad D^{s+1}(y) = \int_0 yD^s(z)dz \quad s = 1, 2, ...$$

Thus, the function D^1 is the c.d.f of the distribution under study, $D^2(y)$ is the integral of D^1 from 0 to y, $D^3(y)$ is the integral of D^2 from 0 to y, and so on. By definition, distribution B dominates A at order s if $D^s_A(y) \geq D^s_B(y)$ for all arguments y.

We can see from Fig. 4.15 that for $h = 3, 6$ and 12, SSA forecasting errors are stochastically much smaller than the errors of the other two methods. In addition, it can be seen that the ARIMA forecast errors are slightly smaller than the Holt-Winter forecast errors. In the case of $h = 1$ there is no evident prevalence of any method.

Direction of change predictions

As another measure of forecast accuracy, in addition to RMSE, we also compute the percentage of forecasts that correctly predict the direction of change (for more details see appendix A).

Table 4.7 provides the percentage of forecasts that correctly predict the direction of change, at $h = 1, 3, 6$ and 12 horizons. It also shows whether they are significantly greater than the pure chance ($p = 0.50$). The symbols $*$ and $**$ in the table indicate the 5% and 1% levels of significance. A set of summary results is also given at the bottom of the table. The summary statistics are the average of correct signs for all eight series at $h = 1, 3, 6$ and 12 horizons and overall average for the

three countries. The percentage of correct signs are generally better than those reported in [67]. This is due to the fact that the results for directional change are particulary sensitive to structural change in the out-of-sample period. The percentage of correct signs can be extremely high or low for all three methods depending on whether there is a structural change in the series in the out-of-sample period. The overall percentage of correct signs for SSA are 90%, 91%, 92% and 85% at $h = 1, 3, 6$ and 12 respectively. For the Holt-Winter, these figures are 89%, 91%, 90% and 82%, respectively, which are slightly lower than the SSA. ARIMA models have produced slightly better results (91% and 92%) at horizons $h = 1$ and $h = 3$ but they are lower (90% and 81%) at $h = 6$ and 12 horizons. For all 96 cases (3 countries, 8 series, $h = 1, 3, 6$ and 12 horizons) SSA has produced 93 significant cases at the 1% and 5% level. Similar results were obtained with the Holt-Winter and ARIMA models, giving 93 and 90 significant cases respectively.

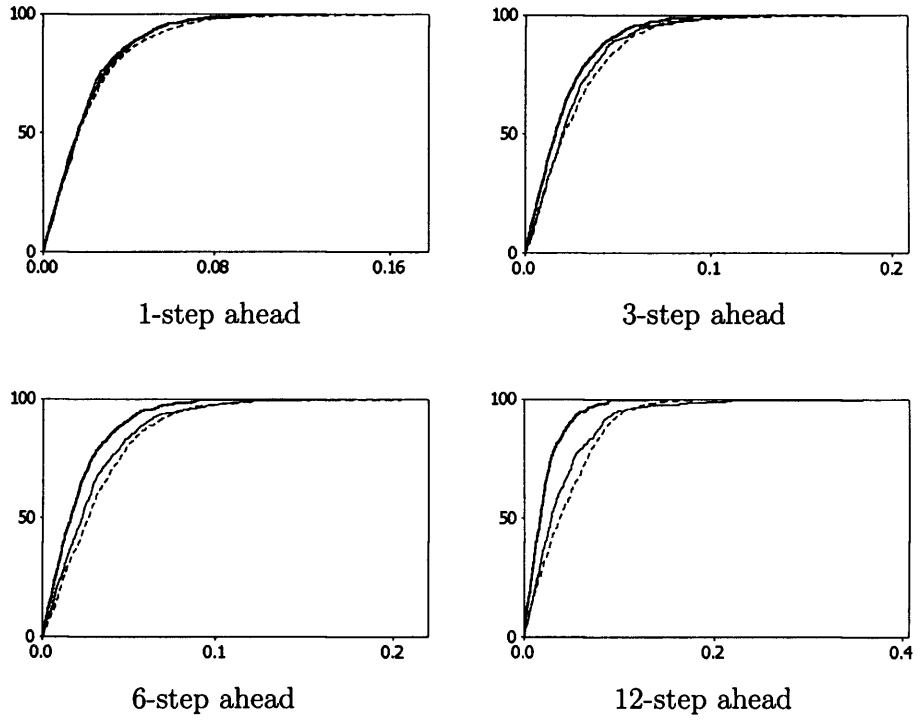


Figure 4.15. The cumulative distribution functions of the absolute values of the out-of-sample errors (for all eight series and 3 countries) obtained by SSA (thick line), ARIMA (thin line) and Holt-Winter (dashed line)

Series	Parameters L	r	$(p,d,q)(P,D,Q)_s$	In-sample: RMSE ARIMA	H-W	SSA	In-sample: RRMSE $\frac{SSA}{ARIMA}$	$\frac{SSA}{H-W}$	h	Out-of-sample: RRMSE $\frac{SSA}{ARIMA}$	$\frac{SSA}{H-W}$	$\frac{BSSA}{ARIMA}$	$\frac{BSSA}{H-W}$
da15	36	1-14	(1,0,0)(0,1,1)	0.012	0.010	0.007	0.58	0.70	1	0.90^{+}	0.78^{*+}	0.93	0.80^{*+}
									3	0.83^{+}	0.79^{+}	0.92^{+}	0.88^{+}
									6	0.77^{+}	0.63^{*+}	0.84^{+}	0.69^{*+}
									12	0.21^{*+}	0.95	0.23^{*+}	1.04
dg24	36	1-14	(0,1,1)(0,1,1)	0.019	0.015	0.009	0.47	0.60	1	0.87	0.77^{*+}	0.93	0.83^{+}
									3	0.65^{*+}	0.67^{*+}	0.70^{*+}	0.71^{*+}
									6	0.58^{*}	0.57^{*+}	0.61^{*}	0.59^{*+}
									12	0.74^{+}	0.80^{+}	0.77^{+}	0.83^{+}
dj27	24	1-16	(0,1,1)(0,1,1)	0.034	0.028	0.005	0.15	0.18	1	0.96	0.90^{+}	0.91	0.85^{*+}
									3	0.81^{+}	0.79^{+}	0.90^{+}	0.89^{+}
									6	0.92	0.92	1.07	1.07
									12	0.30^{*+}	0.80	0.34^{*+}	0.92
dj28	36	1-10	(1,0,0)(1,1,0)	0.026	0.020	0.019	0.73	0.95	1	0.86^{+}	1.06	0.96	1.18
									3	0.84^{*+}	0.99	1.02	1.21
									6	0.79^{+}	0.81^{+}	0.91	0.94
									12	0.42^{*+}	0.83	0.46^{*+}	0.93
dk29	36	1-9	(0,1,1)(0,1,1)	0.026	0.023	0.021	0.81	0.91	1	1.21	0.83^{+}	1.26	0.87^{+}
									3	0.98	0.76^{*+}	1.04	0.81
									6	0.98	0.59^{*+}	0.93	0.56^{*+}
									12	0.76^{+}	0.48^{*+}	0.82	0.52^{*+}
dl31	36	1-11	(0,1,1)(0,1,0)	0.037	0.025	0.020	0.54	0.80	1	1.30	1.48	1.20	1.37
									3	0.93	1.05	0.89	1.00
									6	0.81	0.76*	0.81	0.75
									12	0.42^{*+}	0.47^{*+}	0.56^{*+}	0.63^{*+}
dm34	60	1-13	(0,1,1)(1,1,0)	0.059	0.046	0.027	0.46	0.59	1	1.00	0.96	1.07	1.02
									3	0.76^{*+}	0.80^{*+}	0.83^{+}	0.87^{+}
									6	0.67^{*+}	0.73^{*+}	0.81^{+}	0.88^{+}
									12	0.48^{*+}	0.52^{*+}	0.64^{*+}	0.69^{*+}
e40	36	1-8	(0,1,1)(0,1,0)	0.035	0.024	0.020	0.57	0.83	1	0.93	0.81^{+}	0.97	0.83^{+}
									3	1.02	0.80^{*+}	1.06	0.84^{+}
									6	0.85^{+}	0.67^{*+}	0.92^{+}	0.72^{*+}
									12	0.65^{*+}	0.42^{*+}	0.67^{*+}	0.43^{*+}
Average				**0.031**	**0.024**	**0.016**	**0.54**	**0.70**	1	**1.00**	**0.95**	**1.03**	**0.97**
									3	**0.85**	**0.83**	**0.92**	**0.90**
									6	**0.80**	**0.71**	**0.86**	**0.78**
									12	**0.50**	**0.66**	**0.57**	**0.75**
Score							8	8	1	5	6	5	5
									3	7	7	5	6
									6	8	8	7	7
									12	8	8	8	7

Table 4.3. Descriptive statistics of Out-of-sample and In-sample errors, UK. $*$ indicates significance for DM test at 10% or less, $+$ indicates significance for encompassing test at 10% or less.

Series	Parameters			In-sample: RMSE			In-sample: RRMSE		h	Out-of-sample: RRMSE			
	L	r	$(p,d,q)(P,D,Q)_s$	ARIMA	H-W	SSA	$\frac{SSA}{ARIMA}$	$\frac{SSA}{H-W}$		$\frac{SSA}{ARIMA}$	$\frac{SSA}{H-W}$	$\frac{BSSA}{ARIMA}$	$\frac{BSSA}{H-W}$
da15	60	1-12	(0,1,1)(0,1,1)	0.020	0.020	0.016	0.80	0.80	1	0.89	0.89	0.82	0.83
									3	0.69^{*+}	0.62^{*}	0.69^{*+}	0.63^{*}
									6	0.69^{*+}	0.64^{*+}	0.66^{*+}	0.61^{*+}
									12	0.49^{*+}	0.61^{*+}	0.56^{*+}	0.70^{*+}
dg24	120	1-21	(1,1,0)(0,1,1)	0.024	0.023	0.017	0.71	0.74	1	0.89	0.84	0.98	0.97
									3	0.66^{*+}	0.57^{*}	0.78^{+}	0.67^{*}
									6	0.70^{*+}	0.43^{*+}	0.76^{+}	0.47^{*+}
									12	0.57^{*+}	0.31^{*+}	0.66^{*+}	0.36^{*+}
dj27	60	1-19	(0,1,1)(0,1,1)	0.034	0.032	0.019	0.56	0.59	1	1.59	1.45	1.24	1.13
									3	1.25	1.18	1.01	0.95
									6	0.94	0.76^{*+}	0.73^{*+}	0.58^{*+}
									12	0.56^{*+}	0.47^{*+}	0.44^{*+}	0.37^{*+}
dj28	120	1-18	(0,1,1)(0,1,1)	0.028	0.027	0.021	0.75	0.78	1	0.97	0.89	0.87	0.79
									3	0.75^{*}	0.61^{*}	0.74^{*}	0.61^{*}
									6	0.49^{*+}	0.40^{*+}	0.50^{*+}	0.41^{*+}
									12	0.23^{*+}	0.19^{*+}	0.21^{*+}	0.17^{*+}
dk29	48	1-18	(2,1,0)(0,1,1)	0.035	0.033	0.017	0.49	0.52	1	1.49	1.24	1.04	0.87
									3	1.37	1.03	1.00	0.75
									6	1.01	0.74^{*+}	0.78^{*+}	0.57^{*+}
									12	0.65^{*+}	0.47^{*+}	0.52^{*+}	0.38^{*+}
dl31	48	1-18	(0,1,1)(0,1,1)	0.029	0.028	0.015	0.52	0.54	1	1.48	1.41	1.31	1.25
									3	1.17	1.22	1.05	1.09
									6	0.82^{+}	0.79^{+}	0.75^{*+}	0.72^{*+}
									12	0.54^{*}	0.49^{*+}	0.45^{*}	0.42^{*+}
dm34	60	1-18	(0,1,2)(0,1,1)	0.096	0.092	0.064	0.67	0.70	1	0.72^{*+}	0.45^{*+}	0.84^{+}	0.52^{+}
									3	0.73^{*+}	0.41^{*+}	0.79^{+}	0.44^{*+}
									6	0.74^{+}	0.40^{*+}	0.53^{*+}	0.29^{*+}
									12	0.85	0.44^{*+}	0.83	0.43^{*+}
e40	60	1-15	(0,1,1)(0,1,1)	0.029	0.028	0.019	0.66	0.68	1	0.97	0.96	0.94	0.92
									3	0.75^{*+}	0.76^{*+}	0.71^{*+}	0.71^{*+}
									6	0.69^{*+}	0.70^{*+}	0.67^{*+}	0.68^{*+}
									12	0.62^{*+}	0.62^{*+}	0.61^{*+}	0.61^{*+}
Average				**0.037**	**0.035**	**0.023**	**0.65**	**0.67**	1	**1.12**	**1.02**	**1.01**	**0.91**
									3	**0.92**	**0.80**	**0.85**	**0.73**
									6	**0.76**	**0.60**	**0.67**	**0.54**
									12	**0.57**	**0.45**	**0.53**	**0.43**
Score							**8**	**8**	1	**5**	**5**	**5**	**6**
									3	**5**	**5**	**5**	**7**
									6	**7**	**8**	**8**	**8**
									12	**8**	**8**	**8**	**8**

Table 4.4. Descriptive statistics of Out-of-sample and In-sample errors, Germany. $*$ indicates significance for DM test at 10% or less, $+$ indicates significance for encompassing test at 10% or less.

Series	Parameters L	r	$(p,d,q)(P,D,Q)_s$	In-sample: RMSE ARIMA	H-W	SSA	In-sample: RRMSE $\frac{SSA}{ARIMA}$	$\frac{SSA}{H-W}$	h	Out-of-sample: RRMSE $\frac{SSA}{ARIMA}$	$\frac{SSA}{H-W}$	$\frac{BSSA}{ARIMA}$	$\frac{BSSA}{H-W}$
da15	60	1-12	(1,0,0)(0,1,1)	0.024	0.023	0.014	0.58	0.61	1	0.91	0.78	0.78*	0.67*
									3	0.76*+	0.68*+	0.70*+	0.64*+
									6	0.75*+	0.73+	0.71*+	0.69+
									12	0.80*+	0.67*+	0.76*+	0.63*+
dg24	120	1-21	(0,1,1)(0,1,1)	0.028	0.024	0.017	0.61	0.71	1	0.82	0.79*	0.78	0.75*
									3	0.92	0.90	0.90	0.89
									6	0.85	0.81	0.92	0.88
									12	1.01	1.00	1.23	1.15
dj27	60	1-14	(1,1,0)(0,1,1)	0.031	0.029	0.019	0.61	0.66	1	0.93	0.99	0.83	0.89
									3	0.70*	0.74*+	0.67*	0.71*+
									6	0.50*+	0.56*+	0.51*+	0.56*+
									12	0.39*+	0.53*+	0.40*+	0.56*+
dj28	120	1-18	(0,1,3)(1,1,0)	0.029	0.026	0.017	0.59	0.65	1	0.77*	0.62*+	0.80	0.65*+
									3	0.74*	0.57*+	0.76*	0.60*+
									6	0.63	0.48*+	0.66	0.50*+
									12	0.56*+	0.38*+	0.57*+	0.38*+
dk29	48	1-18	(3,1,0)(0,1,1)	0.028	0.029	0.019	0.68	0.66	1	1.06	1.08	0.98	1.01
									3	1.15	1.05	1.08	0.99
									6	1.15	1.03	1.12	1.00
									12	0.98	0.73*+	0.90	0.67*+
dl31	48	1-18	(0,1,1)(0,1,1)	0.034	0.033	0.022	0.65	0.67	1	1.16	1.10	1.19	1.14
									3	1.06	0.99	1.11	1.03
									6	0.82	0.79+	0.83	0.80+
									12	0.61*+	0.67*+	0.70*+	0.76*+
dm34	60	1-18	(0,1,1)(0,1,0)	0.081	0.077	0.074	0.91	0.96	1	0.94	1.01	0.84	0.90
									3	0.80	0.91	0.75	0.82
									6	0.65*+	0.81+	0.60*+	0.75+
									12	0.42*+	0.57*+	0.40*+	0.55*+
e40	60	1-15	(0,0,8)(1,1,0)	0.048	0.037	0.018	0.38	0.49	1	0.93+	0.86*+	0.87+	0.80*+
									3	0.75*+	0.78*+	0.69*+	0.71*+
									6	0.65*	0.75*+	0.58*+	0.68*+
									12	0.68*	0.71*	0.63*	0.66*
Average				**0.038**	**0.035**	**0.025**	**0.63**	**0.68**	1	**0.94**	**0.90**	**0.88**	**0.85**
									3	**0.86**	**0.83**	**0.83**	**0.80**
									6	**0.75**	**0.75**	**0.74**	**0.74**
									12	**0.69**	**0.66**	**0.70**	**0.67**
Score							**8**	**8**	1	**6**	**5**	**7**	**6**
									3	**6**	**7**	**6**	**7**
									6	**7**	**7**	**7**	**7**
									12	**7**	**7**	**7**	**7**

Table 4.5. Descriptive statistics of Out-of-sample and In-sample errors, France. $*$ indicates significance for DM test at 10% or less, $+$ indicates significance for encompassing test at 10% or less.

Method	N	Mean	S.D	Min	Median	Max
1-step ahead						
Holt-Winter	720	0.00297	0.03109	-0.13771	0.00440	0.16733
ARIMA	720	0.00014	0.02808	-0.13844	0.00165	0.10497
SSA	720	0.00010	0.02837	-0.08982	-0.00034	0.087198
3-step ahead						
Holt-Winter	672	0.00521	0.03555	-0.15961	0.00728	0.19733
ARIMA	672	0.00085	0.03281	-0.14697	0.00284	0.10402
SSA	672	-0.00025	0.02855	-0.09839	-0.00069	0.088908
6-step ahead						
Holt-Winter	600	0.00920	0.04115	-0.18965	0.01150	0.20733
ARIMA	600	0.00347	0.03853	-0.20505	0.00695	0.11062
SSA	600	0.00003	0.02903	-0.13882	0.00063	0.08908
12-step ahead						
Holt-Winter	456	0.01767	0.05278	-0.18090	0.02029	0.14733
ARIMA	456	0.00938	0.05452	-0.35677	0.01424	0.19970
SSA	456	0.00146	0.02952	-0.13039	0.00110	0.09062

Table 4.6. Descriptive statistics of out-of-sample errors.

Series	Holt-Winter				ARIMA				SSA			
	1	3	6	12	1	3	6	12	1	3	6	12
	UK											
Food product	0.87**	0.89**	1.00**	0.89**	0.83**	0.96**	1.00**	0.68	0.90**	0.96**	0.92**	0.74*
Chemicals	0.97**	0.96**	0.92**	0.89**	0.97**	0.93**	0.96**	0.79**	0.97**	0.93**	0.80**	0.89**
Basic metals	0.80**	0.93**	0.76**	0.84**	0.80**	0.86**	0.72*	0.79**	0.73**	0.82**	0.80**	0.74*
Fabricated metal	0.97**	0.93**	0.88**	0.84**	0.93**	0.89**	0.92**	0.84**	0.93**	0.96**	1.00**	0.74*
Machinery	0.90**	0.93**	0.80**	0.74*	1.00**	1.00**	0.96**	0.84**	0.90**	0.93**	1.00**	0.95**
Electrical machinery	0.87**	0.86**	0.84**	0.58	0.93**	0.82**	0.92**	0.53	0.77**	0.89**	0.92**	0.74*
Vehicles	0.90**	0.93**	0.96**	0.84**	0.90**	0.93**	0.96**	0.84**	0.97**	0.79**	0.92**	0.84**
Electricity and gas	0.93**	0.93**	1.00**	0.84**	0.97**	0.96**	0.44	0.89**	1.00**	1.00**	1.00**	0.68
Average	**0.90**	**0.92**	**0.90**	**0.81**	**0.92**	**0.92**	**0.86**	**0.78**	**0.90**	**0.91**	**0.92**	**0.79**
	Germany											
Food product	0.90**	0.78**	0.92**	0.79**	0.90**	0.75**	0.88**	0.84**	0.93**	0.86**	0.92**	0.95**
Chemicals	0.86**	0.89**	0.72*	0.79**	0.87**	0.89**	0.92**	0.89**	0.87**	0.93**	0.92**	1.00**
Basic metals	0.83**	0.79**	0.84**	0.63	0.87**	0.82**	0.84**	0.68	0.80**	0.75**	0.88**	0.89**
Fabricated metal	0.87**	0.93**	0.88**	0.63	0.90**	0.93**	0.88**	0.63	0.77**	0.96**	1.00**	1.00**
Machinery	0.97**	0.96**	0.92**	0.79**	0.97**	0.96**	0.96**	0.84**	0.90**	0.89**	0.88**	1.00**
Electrical machinery	0.90**	0.93**	0.96**	0.89**	0.90**	0.96**	0.96**	0.89**	0.83**	0.86**	0.96**	1.00**
Vehicles	0.80**	0.75**	0.88**	0.58	0.87**	0.89**	0.92**	0.79**	0.90**	0.86**	0.96**	0.95**
Electricity and gas	0.93**	0.93**	1.00**	0.84**	0.97**	0.89**	1.00**	0.84**	0.90**	0.93**	0.92**	0.68
Average	**0.88**	**0.87**	**0.89**	**0.74**	**0.90**	**0.89**	**0.92**	**0.80**	**0.86**	**0.88**	**0.93**	**0.93**
	France											
Food product	0.90**	0.93**	0.92**	0.84**	0.93**	1.00**	0.92**	0.95**	0.93**	0.93**	1.00**	0.79**
Chemicals	0.90**	1.00**	0.88**	0.95**	0.90**	1.00**	0.92**	0.95**	0.93**	0.93**	0.76**	0.95**
Basic metals	1.00**	0.86**	0.88**	0.95**	1.00**	0.89**	0.80**	0.89**	1.00**	0.96**	1.00**	0.89**
Fabricated metal	0.97**	0.93**	0.92**	1.00**	0.93**	1.00**	1.00**	0.95**	0.97**	0.96**	1.00**	0.95**
Machinery	0.93**	1.00**	0.96**	0.95**	0.90**	1.00**	0.96**	1.00**	0.97**	0.86**	0.80**	0.89**
Electrical machinery	0.83**	0.86**	0.84**	0.89**	0.87**	0.89**	0.84**	0.84**	0.97**	0.93**	0.88**	0.89**
Vehicles	0.93**	0.96**	0.84**	0.84**	0.87**	0.89**	0.84**	0.63	0.87**	0.93**	0.80**	0.84**
Electricity and gas	0.77**	0.96**	1.00**	0.89**	0.87**	0.96**	1.00**	0.89**	0.87**	0.96**	1.00**	0.53
Average	**0.90**	**0.94**	**0.91**	**0.91**	**0.91**	**0.96**	**0.91**	**0.89**	**0.94**	**0.93**	**0.91**	**0.84**
Overall Average	**0.89**	**0.91**	**0.90**	**0.82**	**0.91**	**0.92**	**0.90**	**0.81**	**0.90**	**0.91**	**0.92**	**0.85**

Table 4.7. Out-of-sample percentage of forecasts of correct sign. * indicates significance at 5% and ** indicates significance at 1%.

4.3 Iranian National Account Time Series

Econometric methods have been widely used to forecast the evolution of quarterly and yearly national account data sets. However, many of these structural or time series forecasting models have failed to accurately predict the growth rate of Gross Domestic Product (GDP) or the turning points of business cycles in the industrial economies (see, for example, [71]).

Many factors could affect the national economies and hence the national account data which are at best inaccurate representation of the macroeconomic variables because of measurement noise. The exogenous factors that cause instability in macroeconomies including technological changes, government policy changes, changes in the preferences of the consumers, and other events. These shocks cause structural changes in these time series making them nonstationary. Development of a methodology which is robust under these changes is of paramount importance in accurate prediction of macroeconomic time series.

An important feature of SSA is that it can be used for analyzing relatively short and non-stationary series . In the following we apply the SSA technique to 32 original data sets, 16 quarterly and 16 yearly, which are taken from the Central Bank of the Islamic Republic of Iran (CBI). Hassani and Zhigljavsky [12] used the series of Iranian GDP (quarterly) as the main data set for illustrating details of the practical application of the SSA methodology for short time series.

4.3.1 Analysis of Iranian National Account

In this section we demonstrate the capability of SSA by applying it to the analysis and forecasts of the Iranian national account data (for

a comprehensive analysis see [12]). The data sets describe the main economic features of the Islamic Republic of Iran and is provided on the web-site of the Central Bank of the Islamic Republic of Iran[4]. The sets of data are quarterly and yearly. There are 16 quarterly data sets each containing 68 data points over the period of 1988 to 2004 (measured in billion rails, the official currency of Iran).

These sets of data are: 1 – Agriculture, 2 – Oil and Gas, 3 – Industries and Mines, 4 – Manufacturing, 5 – Mining, 6 – Electricity, Gas and Water Supply, 7 – Construction, 8 – Services, 9 – Trade, Restaurants and Hotels, 10 – Transportation, Warehousing and Communication, 11 – Financial Services, 12 – Real Estate and Professional Services, 13 – Public Service, 14 – Social, Personal and Domestic Services, 15 – Imputed Bank Services Charge and 16 – Gross Domestic Product (GDP) in Basic Price. We shall refer to these data sets as Series 1 to Series 16, respectively. Fig. 4.16 displays Series 1 – 16.

It is customary in econometrics to take the logarithms of the data describing economic features. Therefore, we make a parallel analysis of the data taken in the logarithmic scale. Fig. 4.17 displays Series 1 – 16 in the logarithmic scale (the arrangement of the series is the same as in Fig. 4.16).

We also consider 16 yearly data sets which contain 45 observations each covering the period of 1959 to 2003 (measured in billion rails). These data describe exactly the same economic features as Series 1–16. We shall refer to these data as Series 17 – Series 32. Fig. 4.18 displays these series. Fig. 4.19 displays Series 17 – 32 in the logarithmic scale.

On the website of central bank of Iran one can find the Iranian na-

[4]www.cbi.ir

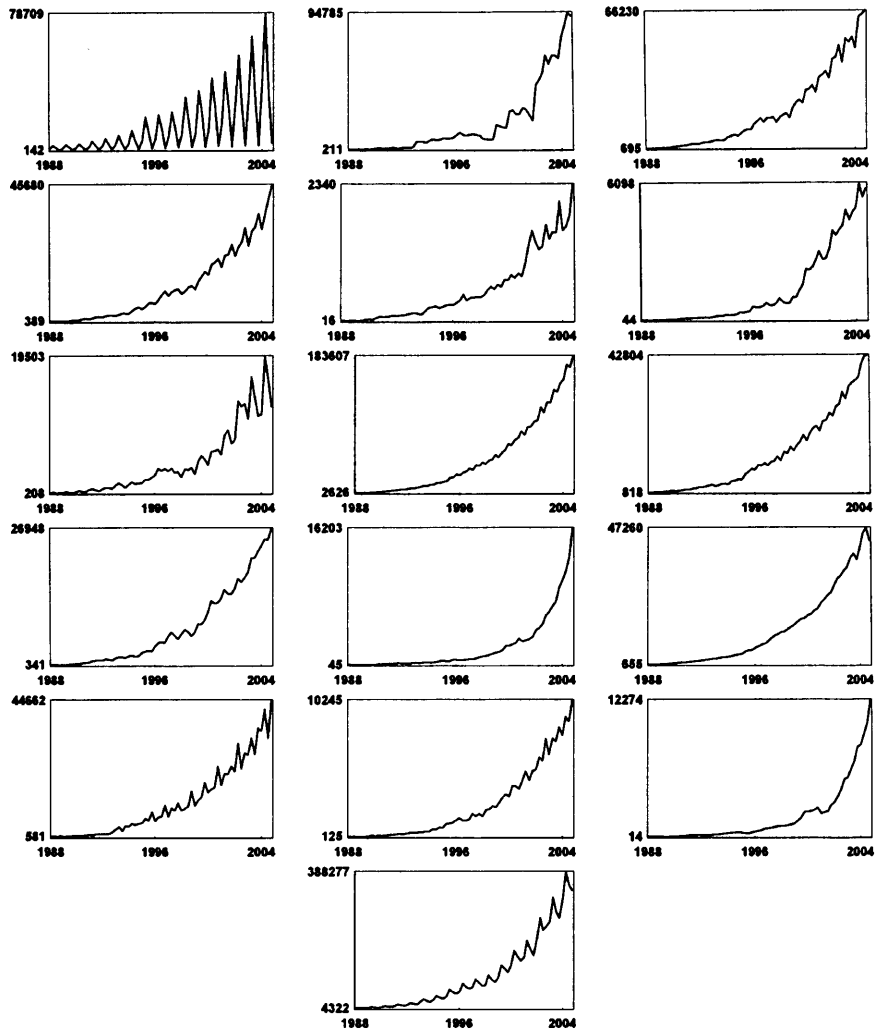


Figure 4.16. Series 1–16.

tional accounts quarterly data adjusted to seasonal effects. However, we use the original, non-adjusted data since one of our aims is to illustrate the capability of the SSA technique for extracting trend and oscillations from the data. We then use the approximated trend and oscillations for forecasting the data.

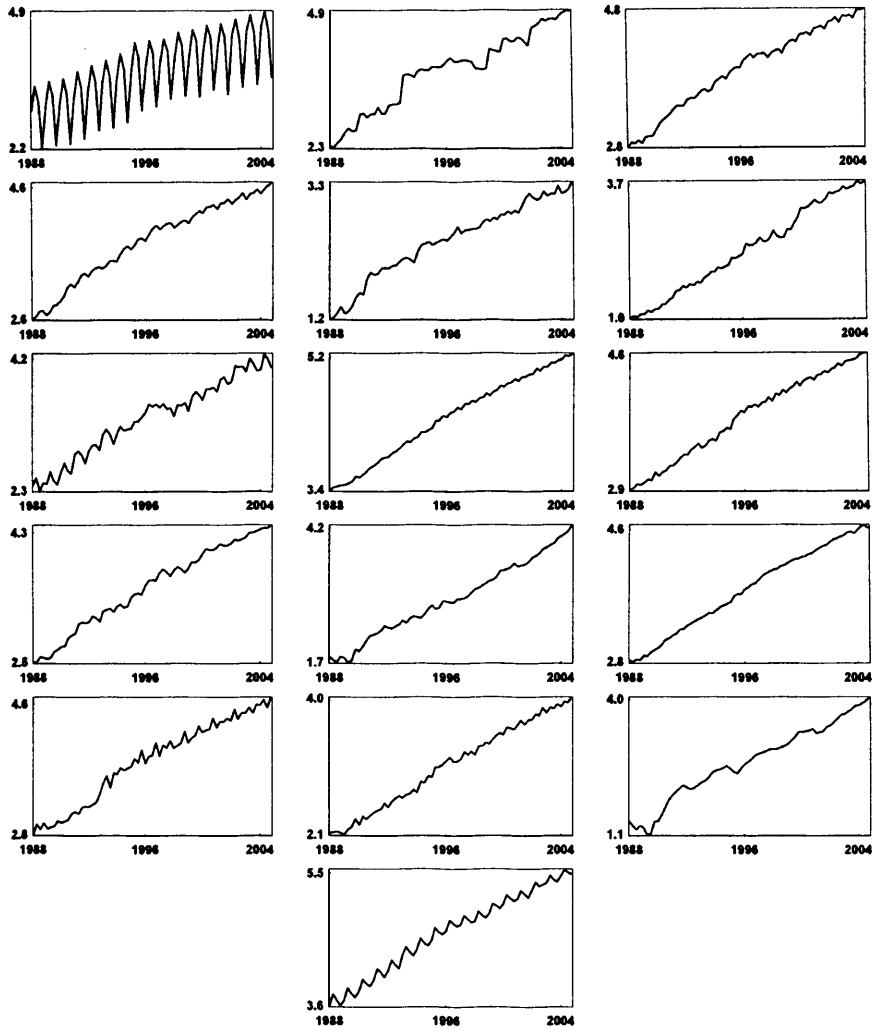


Figure 4.17. Series 1–16 in the logarithmic scale.

4.3.2 Analysis of quarterly data sets

For each series, we have performed SSA analysis and forecast. We have removed the last four points of each series (Q1 – Q4 of 2004), made an SSA approximation for the period 1988 to 2003 and forecasted the data for the four quarters of 2004. In each analysis, we choose the SSA parameters (which are the window length and the number of eigentriples chosen for approximation) to optimize the approximation of the series keeping the window length L large enough.

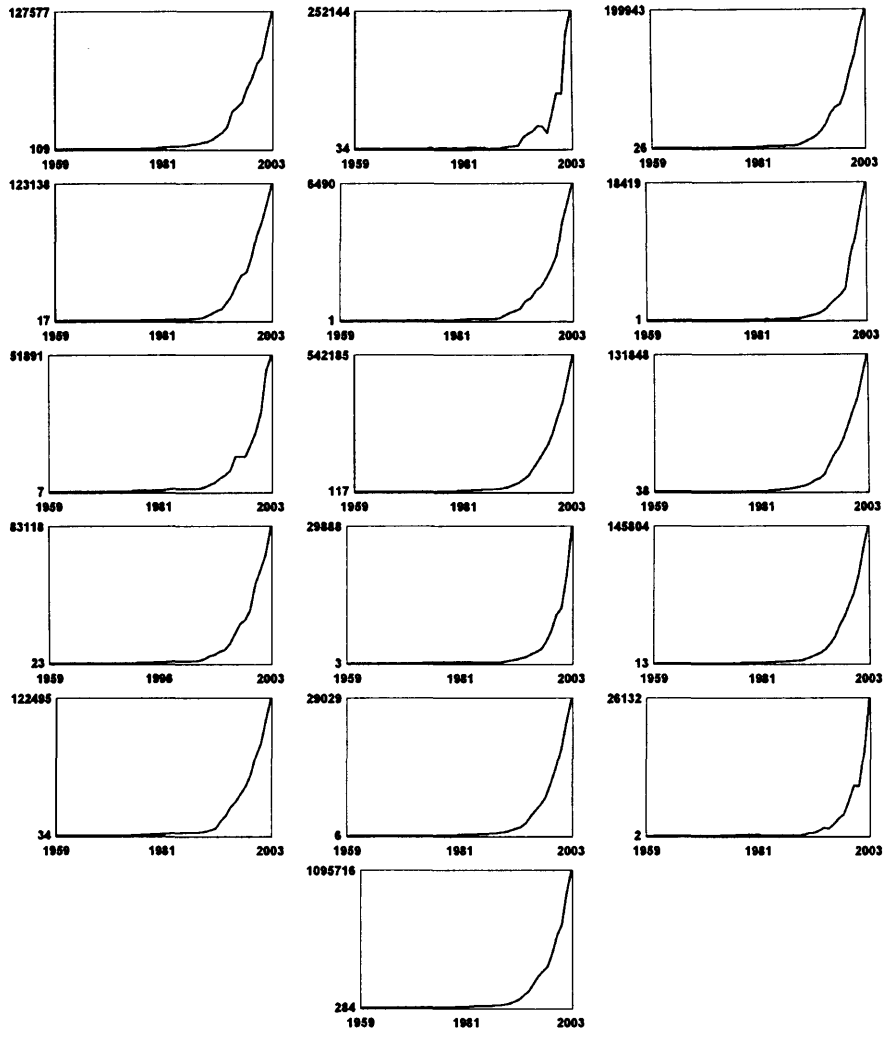


Figure 4.18. Series 17–32.

For each forecasted value (Q1 – Q4 of 2004), we have computed the relative error of the forecast (in percent). To summarize the quality of the forecast, we provide the MARE which is simply the average of the four absolute relative errors (in percent) for each series.

In parallel, we have performed SSA analysis and forecast for the data taken in the logarithmic scale. All the corresponding results are presented in Table 3 (in brackets). When the SSA analysis was performed in the log-scale, for computing the relative error of the forecast,

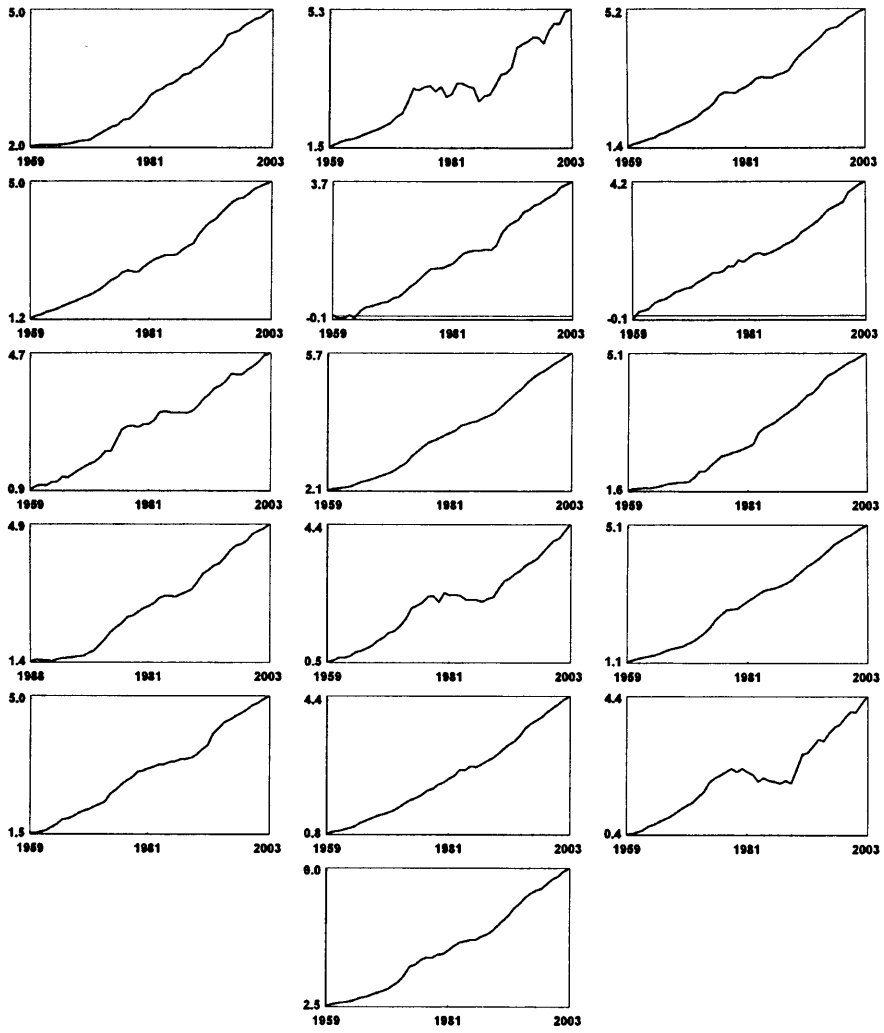


Figure 4.19. Series 17–32 in the logarithmic scale.

we have transformed the forecasted data back to the original scale. We needed to do this in order to be able to compare these results with the results of the original analysis.

Table 4.8 shows the results. Columns 2 and 3 show the parameters of the SSA algorithm (the window length L, see Stage 1 of the SSA algorithm, and the eigentriples chosen r, see Stage 3). Note that using this information and the SSA-Caterpillar software [18], anyone can repeat the results presented in the table).

In each cell in columns 4–7, there are two numbers: the first one is the relative error of the forecast (in percent) for the original series for a given quarter of 2004 and the second one (in brackets) is the value of the relative error of the corresponding forecast when the analysis was performed after taking the logarithms of the series. In the last column, the bold font indicates the lower of the two values. Table 4.8 clearly demonstrates that taking logarithms of the data does not improve the quality of the SSA forecast (on the opposite, it typically leads to its deterioration). This is related to the fact that the quarterly data have periodic components which are easier to extract when the data are considered in the original scale (taking logarithms produces additional smoothing and makes extraction of periodic components more difficult).

We consider the SSA forecasts for all 16 series as very good (an exception is Series 5 and partly Series 7 and 11). The success of the analysis means that in most cases, SSA was able to approximate both the trends and the periodic components with high accuracy. Of course, this is also related to the fact that the economy of Iran was developed steadily during the period 1988 – 2004 (the Iran-Iraq War ended in 1988).

Ser.	L	r	Relative Absolute Error%				MRAE %
			Q1	Q2	Q3	Q4	
1	17 (32)	1-4 (1-10)	3.55 (6.03)	0.87 (2.50)	4.32 (4.50)	0.81 (12.60)	**2.39** (6.41)
2	32 (5)	1,6-7 (1)	0.06 (2.23)	1.24 (.021)	5.77 (0.62)	4.32 (17.0)	**2.99** (5.05)
3	32 (32)	1-7 (1-7)	1.47 (0.35)	2.16 (1.54)	0.98 (0.68)	5.38 (17.4)	**2.50** (5.01)
4	32 (32)	1-7 (1-7)	2.06 (1.66)	0.73 (4.03)	7.17 (9.56)	1.78 (4.22)	**2.93** (4.85)
5	12 (12)	1,2 (1,2)	3.75 (19.1)	6.01 (19.3)	2.95 (12.2)	13.3 (8.63)	**6.51** (14.8)
6	16 (16)	1,2,4-7 (1-4)	1.96 (0.72)	0.02 (9.04)	2.25 (1.22)	1.92 (4.81)	**1.54** (3.95)
7	32 (8)	1-5 (1-4)	10.4 (15.7)	11.9 (9.02)	1.44 (6.17)	0.34 (5.46)	**6.05** (9.09)
8	32 (32)	1-10 (1-4)	0.44 (1.46)	0.06 (0.25)	0.63 (0.07)	1.10 (5.22)	**0.56** (1.75)
9	32 (32)	1-5 (1-5)	1.02 (3.23)	0.74 (3.75)	4.58 (5.83)	0.60 (3.72)	**1.74** (4.13)
10	32 (32)	1,4-7(1-3)	1.35 (1.24)	0.00 (0.71)	4.78 (6.39)	0.22 (2.20)	**1.59** (2.63)
11	5 (10)	1,2 (1,2)	0.32 (0.17)	3.95 (3.40)	4.40 (2.12)	5.24 (8.22)	**3.48** (3.65)
12	12 (10)	1-4,6 (1,2)	4.29 (0.54)	0.30 (5.88)	0.55 (5.84)	2.01 (8.92)	**1.79** (5.20)
13	32 (32)	1-7 (1-5)	0.77 (2.84)	2.69 (3.04)	1.56 (8.67)	2.21 (1.13)	**1.79** (3.92)
14	32 (32)	1-5 (1-5)	4.10 (2.15)	2.83 (0.81)	2.64 (2.31)	1.29 (0.57)	2.72 **(1.46)**
15	8 (5)	1 (1)	1.12 (6.34)	1.04 (2.29)	0.60 (3.30)	8.17 (0.22)	**2.73** (3.11)
16	32 (24)	1-4 (1-4)	0.91 (0.55)	1.88 (0.03)	0.42 (6.49)	0.06 (0.15)	**0.82** (1.81)

Table 4.8. Relative Absolute Error and Mean Relative Absolute Error for Series 1 – 16 before and after taking the logarithm.

Ser.	L	r	Relative Absolute Error%			MRAE %
			2001-2	2002-3	2003-4	
17	5 (5)	1,2 (1,2)	8.59 (8.48)	1.13 (1.58)	0.86(0.28)	3.52 (**3.45**)
18	3 (12)	1 (1)	24.8 (17.5)	15.9 (19.5)	2.06 (1.51)	14.2 (**12.8**)
19	7 (5)	1,2 (1,2)	1.43 (5.75)	1.64 (6.35)	6.84 (2.35)	**3.30** (4.05)
20	5 (5)	1,2 (1,2)	11.3 (0.12)	3.80 (3.42)	4.83 (6.80)	6.66 (**3.45**)
21	7 (11)	1 (1,2)	3.13 (5.25)	1.50 (1.00)	7.34 (3.20)	3.99 (**3.15**)
22	21 (9)	1 (1-4)	3.77 (4.51)	3.25 (2.68)	20.2 (3.01)	9.09 (**3.40**)
23	21 (3)	1-3 (1,2)	15.1 (0.08)	5.85 (13.8)	3.16 (0.25)	8.04 (**4.71**)
24	6 (4)	1,2 (1,2)	2.25 (0.00)	0.13 (3.64)	0.32 (7.11)	**0.90** (3.58)
25	5 (3)	1,2 (1,2)	3.17 (2.87)	0.98 (0.15)	0.56 (1.16)	1.57 (**1.39**)
26	4 (14)	1,2 (1-5)	13.7 (0.40)	4.18 (3.86)	2.80 (5.51)	6.92 (**3.25**)
27	12 (9)	1,2 (1,2)	4.33 (11.0)	2.44 (1.22)	6.31 (5.27)	**4.36** (5.84)
28	3 (6)	1 (1,2)	1.65 (4.29)	1.18 (1.05)	3.47 (3.56)	**2.31** (2.97)
29	21 (6)	1,2 (1,2)	2.38 (1.90)	1.66 (0.18)	6.31 (2.30)	**3.45** (1.46)
30	10 (10)	1 (1-3)	1.43 (0.60)	1.32 (2.42)	7.55 (5.23)	3.43 (**2.75**)
31	21 (15)	1-5,7 (1)	16.5 (19.6)	2.35 (0.92)	9.27 (16.6)	**9.38** (12.4)
32	11 (11)	1 (1-3)	0.27 (0.59)	7.20 (8.61)	0.96 (82.16)	**2.81** (3.78)

Table 4.9. The RAE and MRAE for Series 17 – 32 before and after taking the logarithm.

4.3.3 Yearly data sets

In this section we show the results of the application of the SSA technique to 16 yearly data sets (Series 17 –32). These data sets cover the period 1959 to 2003. These series contain 45 points and are shorter than the quarterly series. Moreover, the economic features exhibit clear non-stationary behaviour in this period and therefore it is much much harder to forecast the yearly series than the quarterly series.

We cut off the last 3 years of each series and forecast it to consider the precision of the technique (that is, we will forecast the values for 2001–2003). Here we do not have seasonal components so we only need to extract the trend of these data sets.

Table 4.9 shows the parameters of the SSA algorithm and the results of the forecasts (the structure of this table is the same as that of Table 4.8). The forecast results for the yearly data are generally worse than that for the quarterly data sets. The main reason for this is the fact

that during the period 1959 to 2003 there were significant changes in the dynamics of the Iranian economic features, see Fig. 4.18 and especially Fig. 4.19. These changes can be associated with the start and the end of the Iran-Iraq War (1980 – 1988). Note that the changes can easily be detected by SSA, see [13] for information about using SSA for detection of changes in time series.

One may note from Table 4.9, that contrary to the case of the quarterly data, the forecast based on the analysis of the series in the logarithmic scale often gives better results. This is perhaps related to the fact that the yearly series do not have seasonal components which are easier to extract when the data is in the original scale.

4.3.4 Iranian Inflation rate series

Next, we present the forecasting results for inflation rate based on the monthly Iranian Consumer Price Index (CPI) series for the short and long horizons $h = 1, 3, 6$ and 12. In fact, we used monthly CPI data for the period Mar. 1990 - Sep. 2007. We used Jan. 1990 to Aug. 2004 CPI observations as training set and Sep. 2004 to Sep. 2007 observations for out-of-sample prediction. We select the window length $L = 60$ and the first 19 eigenvalues for reconstructing the original series and consider remaining eigentriples (20–60) as noise for forecasting inflation rate based on the CPI price index over period Sep. 2004 to Sep. 2007. We also use the RW model as a benchmark model in the comparative analyses. The use of the random walk model as a benchmark model should not imply that we believe the model is an optimal forecasting method. We use this model because it is a naive model. The point here is that a superior performance of random walk model (RW) would ren-

der the analyst's method useless. As a measure of prediction accuracy, here we use RMSE. If RMSE < 1, then SSA procedure outperforms RW model.

Fig. 4.21 shows the CPI series and also inflation rate series based on the CPI series. Visual analysis of Fig. 4.21 indicates that the CPI series has a trend and this trend can be approximated by a function increasing exponentially fast. A harmonic seasonal component with decreasing amplitude is also clearly seen in Inflation rate series. In the following, we only consider Inflation rate series.

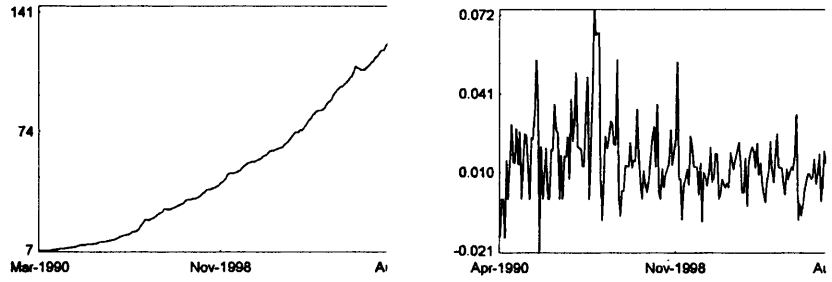


Figure 4.20. CPI series (left) and inflation rate series based on the CPI series (right) Mar. 1990 - Sep. 2007.

Table 4.19 shows the RMSEs for SSA/random walk for h-step ahead forecasts of inflation rate based on the CPI series for N forecasted data points. Without exception, SSA outperforms the random walk predictions in all h-step ahead forecasts. In fact, SSA method is up to 27% more efficient compared to the RW model. Table 4.19 also presents the results of Diebold and Mariano test indicating whether the discrepancies between SSA and RW model forecasting procedures are statistically significant. ** and * imply significance at 1% and 10% confidence levels, respectively. The results of this table confirm that, for all cases, the differences are significant at 1% confidence level.

Additionally, Table 4.19 presents test results for the null hypothesis

of whether the percentages of the direction of changes (DC) are greater than the pure chance (50%). The table shows that all results are statistically significant at 1% and 10% confidence levels. The results of this table also show that MSSA predicts direction of change for 12-step as accurately as it can predict 1-step ahead.

Fig. 4.21 (left) shows the Iranian GDP deflator series (yearly); the data are taken from http://data.un.org. One can see that this series looks very similar to the GDP series. SSA analysis and forecasting results for these two series are also very similar (the results of SSA analysis for the GDP deflator series are not reported here).

Fig. 4.21 (right) shows the Iranian GDP series normalized to the Iranian GDP deflator. The results of SSA forecasting (not reported here) show that it is generally more advantageous to analyze and forecast the two series (namely, Iranian GDP series and Iranian GDP deflator series) separately and then compute the ratio of the forecasts rather than to analyze and forecast the ratio only.

$h=1$			$h=3$			$h=6$			$h=12$		
N	RMSE	DC	N	RMSE	DC	N	RMSE	DC	N	RMSE	DC
36	0.81**	0.69**	34	0.78**	0.68*	31	0.73**	0.74**	25	0.84**	0.67*

Table 4.10. RMSE of the SSA forecast results with respect to the RW method, Diebold-Marino significance test results and direction of change test for inflation rate based on the CPI series.

4.3.5 Forecasting Iranian Macroeconomics series using MSSA

Let us now demonstrate the capability of MSSA by applying it in forecasting 6 quarterly data sets introduced above. We shall refer to these data sets as Series (a) to Series (f); (a)- Gross Domestic Product (GDP) in Basic Price, (b)- Social, Personal and Domestic Services, (c)- Trans-

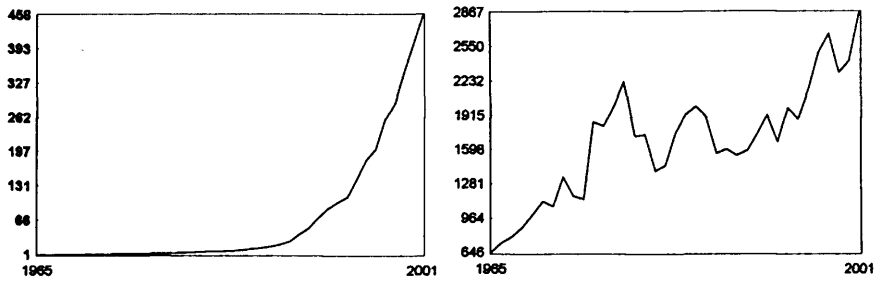


Figure 4.21. Iranian GDP deflator (left side) and Iranian GDP/Iranian GDP deflator (right side).

	L=12		L=16		L=20		L=24		L=28	
Ser.	SSA	MSSA	SSA	MSSA	SSA	MSSA	SSA	MSSA	SSA	MSSA
1	2.110	0.384	0.756	0.386	0.686	0.417	0.774	0.413	0.819	0.477
2	2.487	0.507	1.607	0.522	1.066	0.523	1.285	0.502	1.167	0.547
3	1.528	1.044	1.263	1.082	1.329	1.068	1.354	1.168	1.384	1.267
4	0.927	0.720	0.947	0.743	0.586	0.724	0.450	0.841	0.540	0.998
5	1.338	0.834	0.843	0.857	0.869	0.848	1.434	1.070	1.471	1.199
6	1.195	1.695	1.246	1.693	1.328	1.672	1.111	1.843	0.884	1.864

Table 4.11. MSSA against SSA.

portation, Warehousing and Communication, (d)- Services, (e)- Industries and Mines, and (f)- Public Service. Table 4.11 shows the results for different values of L. As appears from the results, having information of other series helps us to improve the forecasting performance for the series (a)-(e), but it does not help for series (f). The results also indicate that different values of L yields different performance. Table 4.12 shows the MSSA results for a different combination of the series. As the results confirm, choosing a different combination gives different results. The general conclusion is that the MSSA forecasting results are better than the results obtained by SSA if we choose a proper group of series in multivariate approach.

Series	MSSA/SSA				
	L=12	L=16	L=20	L=24	L=28
1–2	1.288	1.205	0.705	0.691	0.681
1–3	0.244	0.734	0.792	0.721	0.683
1–4	0.374	1.204	1.239	1.247	1.263
1–5	0.290	0.641	0.755	0.729	0.721
1–6	0.350	0.900	0.912	0.750	0.673
2–4	0.919	0.396	0.658	0.317	0.329
2–5	0.409	0.376	0.561	0.602	0.614
2–6	0.198	0.297	0.435	0.356	0.384
3–4	0.564	0.772	0.775	0.679	0.601
3–5	1.008	0.731	0.761	0.778	0.751
3–6	0.663	0.801	0.760	0.797	0.852
4–5	0.195	0.347	1.036	1.782	1.557
4–6	0.308	0.349	0.477	0.677	0.568
5–6	0.712	0.675	1.178	1.773	1.497
6–1	0.725	0.632	0.584	0.673	0.796
6–2	1.439	1.398	1.039	0.924	1.141
6–3	1.478	1.3334	1.236	1.505	1.821
6–4	0.971	0.991	0.823	0.774	0.964
6–5	1.118	1.086	1.008	1.139	1.964
1–{3,5}	0.215	0.686	0.754	0.673	0.624
1–{3,5,6}	0.180	0.532	0.608	0.511	0.476
1–{2,3,5,6}	0.164	0.490	0.595	0.513	0.504
1–{2,3,4,5,6}	0.145	0.484	0.594	0.459	0.590

Table 4.12. The MSSA results for different combination.

4.4 Exchange Rate Series

Publication of Meese and Rogoff [72] which showed that a simple random walk model could outperform both linear stochastic time series and structural econometric models in predicting the exchange rates, has generated the voluminous literature of exchange rate economics.

Those financial economists who believe in efficiency of financial markets, however, seriously doubt accurate predictability of the financial asset prices. Efficient Market Hypothesis (EMH) in its weak form implies that the returns of financial asset prices are white noise processes consisting of independent, identically distributed random variables. The white noise nature of the returns implies that the series at level follows a random walk model and is unpredictable.

In spite of the popularity of EMH, mostly in the academic circles, a vast literature dealing with predictions of the financial asset prices exits. Reviewing the empirical exchange rate economics literature one could discern two strands of research in the field that closely follow fundamentalist and chartist (technical analyst and its rough counterpart in academia time series analysts) schism that prevails in prediction of equity prices in the stock markets. In the context of exchange rate economics, the fundamentalists believe that the money supply, the price level, national income, interest rates, productivity, and other relevant economic variables determine exchange rates. The chartists (technical and time series analysts), on the other hand, argue that explaining volatility and accurate predictions of the exchange rates by economic fundamentals is at best futile. They reason that, in spite of daily variations of the exchange rates, the fundamental economic variables seldom, if at all, change in the very short run, making the fundamentals un-

likely explanatory variables, at least, in the short-run. Accordingly, the time series analysts (chartists) attempt to use historical prices of currencies to unravel the underlying dynamics of the exchange rates, and by modeling the dynamics predict future evolutions of the data generating processes of these currencies [73][5].

The most prominent models used in predicting the exchange rates in the fundamentalist tradition include the purchasing power parity theory [5,74,75], sticky-price monetary model [74], the Balassa-Samuleson productivity differential model, the behavioral equilibrium exchange rate model, and the interest rate parity model [76].

Time series analyses of exchange rates, both linear and nonlinear, attempt to predict the exchange rates by using the historical data of interest and without considering the fundamental economic variables that economic theory purports to cause the exchange rate behaviors.

The earlier empirical works in the latter strand of exchange rate economics often used linear stochastic models such as ARIMA process, however, recent development in nonlinear dynamical systems theory, methods of time-delay embedding, and phase space reconstruction has opened up the possibility of testing for presence of nonlinear, deterministic structure in the dynamics of the exchange rates. For example, Soofi and Cao [77], Soofi and Galka [78], and Cao and Soofi [4] and references therein are attempts in prediction and understanding the underlying dynamics of the exchange rates using methods and algorithms from dynamical systems theories that are rarely used in the

[5]Our association of chartists and time series analysts should not be construed that we believe the two approaches use the same set of analytical tools. The association is based on the common belief on the part of the members of the groups that one could use historical data in modeling the dynamics of a set of observations for prediction.

main stream financial economics.

Cheung et al. [79] provides a comprehensive comparative analysis of these competing structural econometric models of exchange rates against a random walk as a benchmark model using quarterly data. The study finds evidence that the structural models outperform the random walk model.

The prediction results based on SSA method are compared with those of a random walk model and the Diebold-Mariano test statistics is used to rule out the comparative results are chance occurrences. Moreover, the direction of change criterion is employed to show the proportion of forecasts that correctly predict the direction of the movement of the series. Finally, to gain a better understanding of prediction accuracy of the methods, the cumulative distribution of the absolute errors of the competing forecasting methods is examined.

The main result of this section is the finding that SSA/MSSA forecasting procedures for exchange rate series are superior to the random walk (RW) forecasts or not. This result may be interpreted from the viewpoint of martingale theory as follows.

A series $\{x_t\}$ is called a martingale (with respect to its own past) if $E_t(x_{t+1}|x_1,\ldots,x_t) = x_t$ for all t. It is widely believed that many financial time series (including exchange rate series) are martingales in this sense. If a series $\{x_t\}$ is a martingale (with respect to its own past), then it is not possible to improve on the random walk (RW) forecast, where $\hat{x}_{t+1} = x_t$ is used as the forecast for x_{t+1}.

The results of this section evaluate an important assumption for prediction of the exchange rate series: indeed, we were unable to build a forecasting method for the exchange rate series that is more precise

than the RW forecast, if the information available was restricted to the series itself. However, when we allowed to use additional information (the values of other exchange rate series of up to time t), then we were able to build a forecast that is superior to the RW forecast. This may imply that the exchange rate time series are not martingales with respect to all available information at the markets. Formally, if $\{x_t\}$ is the series we are interested in and $\{y_t\}$ is a multivariate series of all other currency exchange rates, then our result show that $E_t(x_{t+1}|x_1,\ldots,x_t,y_1,\ldots,y_t) \neq x_t$, which is equivalent to saying that the RW is not the best possible (in the RMSE sense) forecast.

4.4.1 The Data

We shall use two series of daily exchange rates: pound/dollar (UK) and Euro/dollar (EU). We scale each data series according to $y_t \rightarrow y_t / \parallel Y_T \parallel \quad t = 1,\ldots,T$, where $\parallel Y_T \parallel^2 = \sum_{t=1}^{T} y_t^2$. To make sure that all series we are dealing with have the same scale (weight) we adopt the normalization method introduced above.

Fig. 4.22 shows these (rescaled) series over the period 3-Jan-2000 to 8-Dec-2006, in these prediction exercises. Each of these series contains 1810 points. Its very clear that the UK and EU series are highly correlated (indeed, the value of λ between UK and EU series is about 0.77). The value of w-correlation is also about 0.006. It should be mentioned that this correlation only shows the relationship between the main trends of the series.

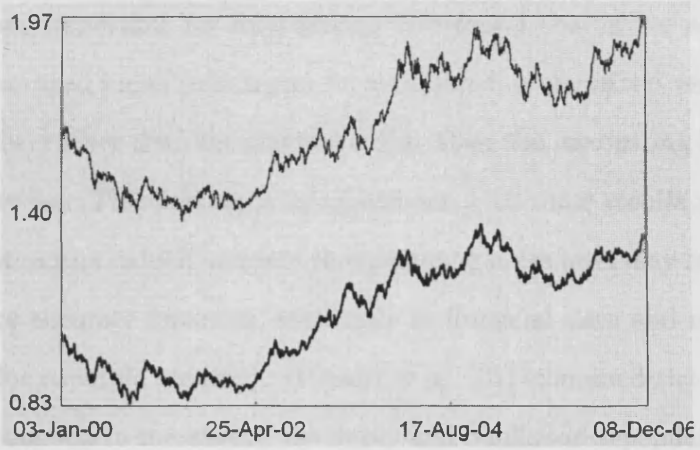


Figure 4.22. The exchange rate series UK (thin line) and EU (thick line) exchange rate series over the period 2000 to 2006.

4.4.2 Trend Analysis

The main discrepancy between SSA and classical time series analysis lies in the notion of trend. For the SSA technique, trend is slowly varying component of the series, which does not contain cyclical / seasonal components. As we do not have obvious periodic components in the series, we only need to extract the trend of these data sets, and for trend extraction, small window length should suffice (for more information about selection of the SSA parameter see [8], chap. 1 and 2).

Fig 4.23 shows the extracted trend of the original series of UK (thin line) and EU (thick line) which are obtained from the first eigentriple and the window length $L = 30$. Note that we can build a more complicated approximation of the trend if we use some other eigentriples and smaller window length. However, the precision we would gain will be very small but the model of the trend will become much more complicated. The value of λ between the trends of UK and EU series is 0.80. We see that the correlation coefficients have slightly increased (in the absolute values). This is due to smoothing. The change is very

small but important for forecasting. We found that if we use bootstrap averaged series (which can be considered as smoothed versions of the series) rather than the original series, then the forecasting becomes more precise. This finding is in agreement with some results reported in the literature, which indicate that reducing noise level may help us to get more accurate forecasts, especially in financial data and nonlinear series (for example see [80]). Hassani et al. [31] examined the effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. They found that noise reduction matters in measuring the linear and nonlinear dependency between two series.

To forecast UK exchange rate series, we shall use rescaled and then bootstrapped EU exchange rate series. Note that we use the original UK series in conjunction with rescaled and bootstrapped EU series.

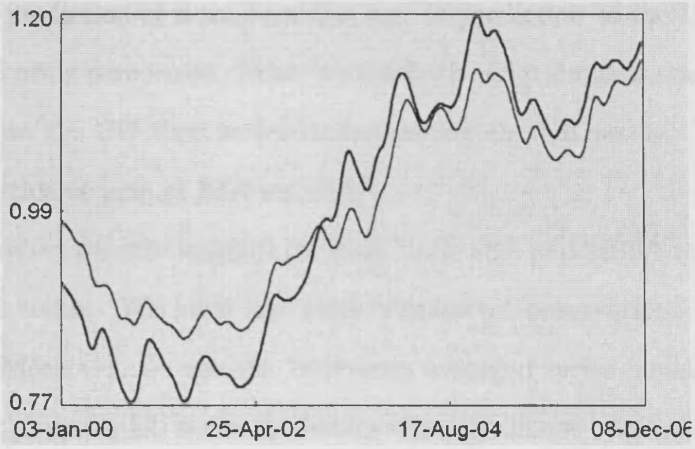


Figure 4.23. Trends of UK (thin line) and EU (thick line) rescaled exchange rate series which are obtained from the first eigentriple.

4.4.3 Results

To acquire the best forecasting accuracy we use different procedures. First, we consider univariate SSA against RW model. Second, we com-

pare MSSA and RW forecasting results. Next, we examine the performance of the MSSA technique with additional information to answer the question whether exchange rate series is martingale. Finally, we use the traditional a-theoretical time series analysis of vector autoregressive, cointgeration, and error correction model to forecast the exchange rate and compare the results with the predictions of SSA. We do not present the results of this latter analysis in this paper because of the error correction model's performance is eminently inferior to those of a RW or SSA models.

To consider the precision of the technique, we forecast all observations of the UK series from 18-Sep-2006 to 8-Dec-2006. We only perform one-step ahead forecasting based on the most up-to-date information available at the time of the forecast. Note that we first use SSA in prediction of a single series, e.g. in prediction of the UK series without using euro series. Next, we use both series simultaneously, that is, we use the EU time series in forecasting the UK series. We shall refer to this version of SSA as MSSA.

We select window length 3 for both Basic SSA and MSSA to forecast the UK series. We have the same number of observations for both series. Moreover, we use the bootstrap averaged series instead of the original series for EU series, to reduce the noise in the original series. It should be noted that if RMSE < 1, then the SSA forecasting procedure outperforms the random walk.

Univariate SSA

In Table 4.13 we represent the results of comparison of RW forecasts with forecasts made by univariate SSA. In the first column we present

the number of forecasting steps. The second column shows the RMSE for each forecasting period. The third and fourth columns show DM and DC statistics, respectively. The last row summarizes the average results. We keep the same procedure for Tables 4.14 and ??.

We have selected 60 data points. The behavior of the series in the chosen period looks very typical. As shown in Fig. 4.22 we have many changes of direction in the series, periods of slow and fast movements of the normalized rates. We observed that the forecast is typically good when there is no sudden radical change of behavior of the series at the forecast point. Alternatively, if there is such a change, the forecast is often misleading.

Overall the results show that Basic SSA perform better than a RW model for the first 30-step ahead observations. However, over a longer horizon, SSA looses its advantage and performs poorly compared to a RW model. Nevertheless, on average for the entire 60-step ahead prediction, SSA has the upper hand, even though it is a marginal advantage. In fact, the forecasting errors are not significantly smaller (in probabilistic sense) than the errors of the RW forecast. The average shows that the SSA forecasts are comparable with the forecasts obtained from a RW model.

We observe that the forecasts obtained from the SSA technique have better performance than RW model in forecasting direction of change. As it can be seen from Table 4.13, the direction of change forecast results using SSA are better than the RW without exception, with 73% accuracy for $N = 60$ increasing to 90% for $N = 10$ compared to 50% for the RW model. The results of the DC test indicate significance at the 1% level.

Therefore, we conclude that we cannot gain substantial improvement in forecasting using univariate SSA. However, the advantage of using SSA is that one can improve the direction of change forecasts. The situation, however, changes drastically when one uses MSSA. In summary, using univariate SSA enables us to improve direction of change at least. Next, we use multivariate version of the SSA technique to improve the accuracy of the forecast.

	SSA		
N	RMSE	DM	DC
10	0.87	-0.46	0.90***
20	0.83	-1.00	0.90***
30	0.95	-0.36	0.80***
40	1.02	0.16	0.75***
50	1.05	0.40	0.72***
60	1.04	0.38	0.73***
Average	0.96		0.80

Table 4.13. Summary of the results for forecasting of UK exchange rate series with SSA and RW. *** indicates the significant results on the 1% level.

Let us now consider a reason on why SSA, on average, performs better than a RW model in forecasting and direction of change prediction. It is well known that the existence of a significant noise level reduces the efficiency of the methods to analyze and model the time series. Two approaches to model the noisy series exist. According to the first one, which is used in classical modeling, one neglects presence of the noise in the series and model the noisy series. According to the second approach, which we use in SSA, we start with filtering the noisy time series in order to reduce the noise level and then model the series. Accordingly, one finds the results by the second approach more effective than the first one if we select a proper method for filtering the series (for more information see [77]).

Multivariate SSA

So far we have used univariate SSA in forecasting exchange rate. As mentioned above, the correlation between UK/dollar and EU/dollar exchange rate is high (it is about 0.77). This motivate us to use multivariate version of the technique. On the other hand, the high correlation between the series implies that there might be causal relationship between these exchange rates. It can be observed from Table 4.14 that the difference between MSSA predictions and RW are significant with respect to all chosen criteria. The results confirm with strong evidence that we have improved both accuracy and direction of change of the forecasting results. Again, the results of the DC test indicate significance at the 1% level. Comparing to univariate case, we have improved the accuracy of the forecasting results from only 4% to 20% on average. Therefore, using the information of EU exchange rate enables us to improve our results up to 16% on average.

	MSSA		
N	RMSE	DM	DC
10	0.84	-0.94	0.80***
20	0.73	-1.91*	0.85***
30	0.81	-1.55	0.83***
40	0.84	-1.64*	0.77***
50	0.81	-2.08**	0.76***
60	0.79	-2.45***	0.78***
Average	0.80		0.80

Table 4.14. Summary of the results for forecasting of UK exchange rate series with MSSA, VAR and RW. Symbols *, **, and *** indicate the significant results on the 10%, 5% and 1% levels, respectively.

Let us now examine the empirical cumulative distribution function (CDF) for the absolute errors of the respective methods. In Fig. 4.24 we display CDF for the absolute errors of the MSSA and RW fore-

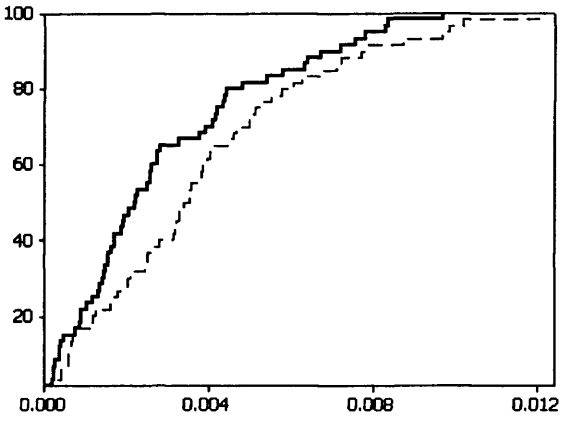


Figure 4.24. Empirical cumulative distribution functions of the absolute errors for MSSA (thick line) and random walk (dashed line).

casts. This plot shows that the empirical distribution of the RW errors stochastically dominates the distribution of the MSSA errors (that is, the RW errors are stochastically larger than the MSSA errors). Note that the Kolmogorov-Smirnov test (the p-value is 0.90), indicates that the distribution of errors for the MSSA forecast does not contradict the hypothesis of normality.

4.4.4 Further Comparisons

In this section we use the traditional econometrics time series in exchange rate predictions. Specifically, given the high correlation between the pound/dollar and EU/dollar exchange rates ($\lambda = 0.77$), we use a 2-variable vector autoregressive (VAR) model (VEC) in exchange rate predicting. This approach to prediction is called a-theoretical, since there is no theoretical justifications in asserting that one exchange rate is a predictor of another one.

The starting point in VAR analysis is testing for presence of unit roots in the time series. We use Augmented Dickey-Fuller method

in testing for presence of unit roots in the exchange rate series. As the unit root test statistics shown in Table 4.15 below indicate, the exchange rates are non-stationary I(1) processes. According to the P-values in Table (4.15), we do not reject the null hypothesis that UK/dollar and EU/dollar exchange rate series, in level with and without trend in model, has a unit root. However, the null hypothesis are rejected in 1th difference which indicates the series are I(1). Therefore, one should use the 1th difference series for further analysis.

Series	Test Statistics	P-value
UK	-0.35	0.91**
EU	-0.34	0.92**

Table 4.15. Augmented Dickey-Fuller test statistics

Next, we test whether a linear combination of the integrated series in the VAR model is stationary, that is, we conduct cointegration test. Using Johansen maximum-likelihood method, we found that the exchange rates are cointegrated series. The results of cointegration test represented in Table 4.16. The results confirm that there is one cointegrating equation at 1% levels. Based on this finding we estimated a error correction model and used it in prediction exercises. The results are decisively inferior to the all models we used in this study, that is, the SSA, MSSA, and RW models.

A question that frequently arises in time series analysis is whether one economic variable can help forecast another economic variable.

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	1% Critical Value
None **	0.0117	22.25	20.04
At most one	0.0005	0.934	6.65

Table 4.16. The results of Cointegration Test. ** denotes rejection of the hypothesis at the 1% level.

Here the question is whether the EU exchange rate can help us in forecasting UK exchange rate series and vice versa. One way to address this question was proposed in [81]. Testing causality, in the Granger sense, involves using F-tests to test whether lagged information on one variable, say X, provides any statistically significant information about another variable, say Y, in the presence of lagged Y. If not, then "Y does not Granger-cause X."

Let us now consider the pairwise Granger Causality Tests for UK and EU exchange rate series. The results have been represented in the following table. As the results show, we would accept that the UK(EU) exchange rate series does Granger Cause EU(UK) exchange rate series as the P-value is smaller that 0.05. In fact we would reject the null hypothesis which is UK(EU) does not Granger Cause EU(UK). These results motivated us to use MSSA.

Null Hypothesis:	F-Statistic	P-value
UK does not Granger Cause EU	6.20674	0.00206
EU does not Granger Cause UK	11.4588	1.1E-05

Table 4.17. The pairwise Granger Causality Tests.

The Granger causality test shows that UK/dollar does Granger cause EU/dollar exchange rate series and vice versa. Therefore, a VAR model can be considered as a benchmark model for multivariate case

rather than RW model. But for consistency with univariate case, we use RW model as a benchmark and also VAR model as another predictive model. Table 4.18 represents results for MSSA and VAR model. It can be observed from Table 4.18 that the VAR model has not good performance in prediction exchange rate model. However, VAR model gives slightly better results for direction of change in comparison to RW model. It improves only 6% in average (from 50% for RW to 56% for VAR). But there is not any improvement in forecasting future data points. In contrast, the difference between MSSA predictions and RW are significant with respect to all chosen criteria. The results confirm with strong evidence that we have improved both accuracy and direction of change of the forecasting results. The results also show that the performance of MSSA is, in average, approximately 20% better than VAR model.

	MSSA			VAR			RRMSE
N	RMSE	DM	DC	RMSE	DM	DC	MSSA/VAR
10	0.84	-0.94	0.80***	0.99	-0.32	0.60	0.85
20	0.73	-1.91*	0.85***	0.98	-0.37	0.70	0.74
30	0.81	-1.55	0.83***	1.02	0.17	0.53	0.79
40	0.84	-1.64*	0.77***	1.02	0.14	0.53	0.82
50	0.81	-2.08**	0.76***	1.03	0.12	0.50	0.79
60	0.79	-2.45***	0.78***	1.03	0.15	0.50	0.77
Average	0.80		0.80	1.01		0.56	0.79

Table 4.18. Summary of the results for forecasting of UK exchange rate series with MSSA/RW, VAR/RW and MSSA/VAR. Symbols *, **, and *** indicate the significant results on the 10%, 5% and 1% levels, respectively.

4.4.5 MSSA results for the Efficient Market Hypothesis

The empirical results of the present study are instructive in examining the efficient market hypothesis controversy. Accordingly, we first present formal discussions of the martingale games, random walk pro-

cesses, their relationship with the EMH, and then we elaborate on the implications of our findings for the EMH.

A stochastic process x_t follows a martingale if

$$E_t(x_{t+1}|\Omega_t) = x_t \tag{4.4.1}$$

where Ω_t is the information set at time t that includes x_t also. Equation (4.4.1) implies that if x_t follows a martingale the best forecast of x_{t+1} is x_t, given the information set Ω_t.

Alternatively, one could present a martingale as a "fair game"– meaning a game that is neither in your favor nor in your opponent's favor– as

$$E_t[(x_{t+1} - x_t)|\Omega_t] = 0 \tag{4.4.2}$$

The implication of the fair game model (4.4.2) in financial economics is that the returns of the asset price x_t are unpredictable, given the information set Ω_t. Accordingly, the information set Ω_t is fully reflected in the asset price, and this is known as the EMH[6].

Note that one may restrict the information set Ω_t only to the asset's past price history, making alternative representation of (4.4.1) and (4.4.2) as

$$E(x_{t+1}|x_t, x_{t-1}, ...) = x_t \tag{4.4.3}$$

or

$$E(x_{t+1} - x_t|x_t, x_{t-1}, ...) = 0 \tag{4.4.4}$$

In the latter representation, again, the EMH suggests that the infor-

[6]We are using EMH in a generic sense, to avoid further discussion of the types of efficient market hypothesis which is not germane to the issue here.

mation contained in the price series of an asset is reflected "instantly, fully, and perpetually" in the asset's current price. Since the price series and the information contained in it are available to all market participants, no one can benefit by attempting to take advantage of the information contained in the price history of an asset by trading in the markets. This reasoning implies that the price movements in the most efficient market are completely random.

A random walk model without drift is represented as follows:

$$x_{t+1} = x_t + \eta_t \tag{4.4.5}$$

where η_t is *i.d.d.*, a white noise process, with zero mean. A random walk model is a martingale, but a more restrictive one, in the sense that it requires both independence of conditional expectation of price changes from the available information (as does the martingale) as well as independence of higher conditional moments (variance, skewness, and kurtosis) of the probability distribution of price changes.

What are the implications of our empirical findings for the EMH? Based on the results of SSA predictions, which were based only on the past price history, we conclude that the currency markets are efficient and follow a random walk process. However, the results based on MSSA which are obtained by including other information, i.e. EU/dollar exchange rate, clearly point to inadequacy of the random walk in modeling exchange rate for predictions. Moreover, the superior results obtained from the direction of change method, also provide additional support for the view that currency markets may not be efficient in the sense discussed above.

4.5 Inflation Rate Series

Accurate prediction of inflation rate has been a subject of great research interest for economists. The keen interest in the subject emerges from pivotally important role accurate prediction of inflation plays in macroeconomic policy analysis and decision making.

Research works on modeling and prediction of inflation began as early as 1950s. In late 1950s Phillips [82] correlated nominal wage inflation with unemployment in the United Kingdom. Inflation studies in the United States modified this model somewhat and searched for a possible relationship between inflation and unemployment rates [83]. In 1960s, Phelps [84] and Friedman [85] both criticized the original Phillips curve analysis by pointing out that these earlier models did not account for the effects of expectations in wage and price determination. These latter analyses led to what is known as the accelerationist Phillips curve that assumes a relationship between the nonaccelerating inflation rate of unemployment (NAIRU) and the output gap [86,87].

Emergence of stagflation in Europe and America in 1970s and breakdown of the inflation-unemployment nexus motivated the theorists to develop 'triangle model' of inflation with the vertices of the triangle consisting of real economic variables (measured by unemployment rate), supply shocks (e.g. energy prices), and inflation 'inertia' (lagged inflation), as well as new-Keynesian Phillips curve [88]. We refer interested reader to [89], for excellent discussions of the development of Phillips curve analysis; and to [90], on empirical estimations of alternative Phillips curve-based models.

Due to failure of the Phillips curve-based model in accurate prediction of inflation rate researchers have used a variety of methods in pre-

dicting inflation rate. These methods include application of dynamics factor models (DFMs) for construction of an index of economic activities as a proxy for unemployment rate for use in the Phillips Curve model, estimation of linear models using financial variables such as interest spreads, stock prices, money supply, among other variables, univariate time series $AR(p)$ as well as $MA(q)$ representations of the inflation data [91,92], and survey techniques [93,94].

In recent years a number of comparative studies of inflation forecasting methods resulted in two major insights about inflation forecasting methods and inflation rate in the United States. First, the studies are inconclusive about the superiority of the competing forecasting methods. For example, Stock and Watson [86] documents that Phillips curve-based models tend to have the most accurate forecast of the inflation in the United States up to 1996. While Atkeson and Ohanian [88] contradicts the conclusion about the relative forecasting accuracy of the Phillips curve-based models and shows that a naive random walk model has a superior predictive capability.

Ang et al. [95] compares four methods of inflation forecasting for the post-1985 and post-1995 periods in the United States, and negates findings by Stock and Watson as well as Atkeson and Ohanian [88] by concluding that the survey-based method tends to outperform the Phillips curve-based model, the term-structure models, and the ARIMA models in inflation forecasting for the United States.

The second insight emerging from the inflation prediction studies [96–98], is that two distinct periods of inflationary pressure are observed in the United States. The first, more volatile period was the period of the early 1970s to mid-1980s. The second, more stable period, the

period of "Great Moderation" in inflation rate as it is known, began in the mid-1980's and has lasted to the present.

This study aims to predict the inflation rate using the United States' consumer price, and chain-weighted GDP indexes. Here the univariate SSA and multivariate SSA are utilized in these predictions which include both the magnitude and direction of changes. Furthermore, out-of-sample predictions are compared with those of alternative methods of inflation prediction, methods such as activity-based NAIRU Philips curve (Atkeson and Ohanian model [88]), $AR(p)$, and random walk as a naive forecasting method.

The RW model is used as a benchmark model in the comparative analyses. The use of the random walk as a benchmark model is motivated by the findings in [91] showing that Atkeson-Ohanian model substantially outperforms the more complicated models in prediction of inflation for the U.S. for 4 and 8 quarters horizons.

The use of the random walk model as a benchmark model should not imply that we believe the model is an optimal forecasting method. We use this model because it is a naive model. The point here is that a superior performance of random walk model would render the analyst's method useless.

Again, we are motivated to use SSA because of its ability in dealing with stationary as well as non-stationary series. Given that the dynamics of the U.S. economy has gone through many variations due to policy and structural changes during the time period under consideration, one needs to make certain that the method of prediction is not sensitive to the dynamical variations. Moreover, contrary to the traditional methods of inflation forecasting (both autoregressive or structural models

that are based on the assumptions of normality and stationarity of the series), SSA is a non-parametric model and makes no prior assumptions about the data.

It should be noted that in some instances, removal of cyclical component would also improve prediction outcomes. For instance, in a related, ongoing study we have discovered that the core CPI series for the United States contains a cyclical component also. The data transformation of the CPI by eliminating both the cyclical and random components of the time series is the main factor contributing to predictive power of SSA method.

The traditional methods for modelling and forecasting time series such as ARMA models suffer from parametric restrictions. For example, in order to optimally fit an ARMA model, the data must be stationary and normally distributed. Although one can transform a non-stationary series by first differencing or de-trending it before fitting an ARMA or ARFIMA (autoregressive fractionally differenced moving average) models to the data, nevertheless, one would loose a great deal of information by such data transformation. These requirements do not exist for SSA, as it does not depend on any parametric model for the trend or oscillations, and does not make any assumptions about the signal or the noise component of the data.

As it was stated above, singular spectrum analysis (or in its multivariate version MSSA) decomposes a time series into its components of trend, cyclical and seasonal variations, as well as noise. Then leaving the noise or cyclical component aside it reconstructs the decomposed series for prediction or for identifying *structural break* or *change point* in the series.

Selection of the window length, L, which in theory of nonlinear dynamics is referred to embedding dimension, is a topic of up most importance in state space reconstruction of observed time series. Such state space reconstruction is required for an understanding of the underlying dynamics of the observed scalar series. However, a discussion of this topic is beyond the scope of the present work. We refer the interested reader to [77] for detail discussions of window length (time delay and embedding dimension in jargon of nonlinear dynamical systems theory) selection. Nevertheless, theory of singular spectrum indicates that the window length $L \leq T/2$ gives a reasonable reconstruction of the dynamical system [8]. However, given the superior predictive performance of the SSA relative to the competing methods, the arbitrary choice is of no practical consequence. Therefore, for brevity sake, we do not apply the usual procedures of determination of time delay and embedding dimension selection in the present study.

4.5.1 Methods used in the previous studies

Phillips Curve and dynamic factor model

The dynamic factor model (DFMs) constructs factors (indexes) as the principal components of the set of predictors consisting of a large number of macroeconomic time series and commodity prices. The index is then used in the Phillips curve-based model as a proxy for the unemployment rate. This approach has been used in [86] and [88], among others.

Specifically, Atkesion and Ohanian [88] estimate the following model

which is a modified version of NAIRU Philips curve model used in [86]:

$$\pi_{t+12}^{12} - \pi_t^{12} = \alpha + \beta(L)x_t + \gamma(L)(\pi_t - \pi_{t-1}) + \eta_{t+12} \qquad (4.5.1)$$

where π_t^{12} is inflation over 12 months as measured by $\pi_t^{12} = 100[\log(p_t) - \log(p_{t-12})]$, p_t denotes the price index in month t. In model (4.5.1), x_t is the activity index constructed using dynamic factor method in conjunction with 158 or 85 monthly time series of the National Economic Activity Index (CFNAI) that is compiled by the Federal Reserve Bank of Chicago. Finally, $\beta(L)$ and $\gamma(L)$ are polynomials in the lag operator L, and η_{t+12} are the error terms and are assumed to be an *iid* series.

Note that the left hand side of (4.5.1) is the difference between the inflation rate of next 12 months and the inflation rate of the last 12 months. Moreover, by letting $\alpha = \beta(L) = \gamma(L) = 0$, we can use (4.5.1) as a random walk process to conduct naive forecasts.

We use the prediction results based on Atkeson-Ohanian model which is the NAIRU-based Phillips curve in this comparative analysis[7].

Autoregressive model

Another approach in inflation forecasting that appears in the literature is modeling the price indexes as $AR(p)$ processes. In this modeling approach Akaike Information or other information criterion in determining the lag order p is often used.

[7]We are grateful to Dr. William Gavin of the Federal Reserve Bank of St. Louis for the generous supply of his inflation prediction data and the predicted data based on Atkeson and Ohanian.

4.5.2 The data

We use several U.S. price indexes in out-of-sample, h-step-ahead moving prediction exercises. These indexes including consumer price index with highly volatile food and energy items (CPI-all), and without highly volatile food and energy items (CPI-core), as well as real-time quarterly chain-weighted GDP price index. Specifically, we used monthly CPI-all and CPI-core data for the period JAN 1986 - DEC 2006. The real-time chain-weighted GDP price series consists of observations starting in the first quarter of 1959 and ending in the third quarter of 1999.

We use sample observations 1959.Q3 to 1991.Q4 of GDP price index for training and observations 1992.Q1 to 1998.Q4 for out-of-sample prediction. Additionally, we used JAN. 1978 to Dec. 1996 CPI observations as training set and Jan. 1997 to Dec. 2006 observations for out-of-sample prediction.

We use moving h-step-ahead prediction, which means that we include all available information for the predictions. This means that for 1-step-ahead prediction, after using $y_1 \cdots y_T$ in prediction of y_{T+1}, we use all observations $y_1 \cdots y_{T+1}$ in prediction of y_{T+2}, and so forth.

In addition to using real-time Chain-weighted GDP price index, we also used the GNP/GDP deflater and we find our prediction results based on these two data sets are very similar.

4.5.3 Forecasting Inflation rate based on the CPI-all and CPI-core series

Next, we present the forecasting results for inflation rate based on the Consumer Price Indices for the long and short horizons. We use MSSA for forecasting inflation rate based on the CPI-all and CPI-core series

over the period Jan-1986 to Dec-1996 that was used as the training set data.

As the first step in using MSSA, we must perform SSA, by choosing the window length L (which is the only parameter in the decomposition stage). Selection of the proper window length depends on preliminary information about the time series. If we know that the time series may have a periodic component with an integer period (for example, if this component is a seasonal component), then to get better separability of this periodic component it is advisable to take the window length proportional to that period. Based on these considerations, we take $L = 60$. The length of CPI-core series is $T = 132$. The value of w-correlation is also about 0.002. Therefore, based on this window length and considering the SVD of the matrix $\mathbf{X}\mathbf{X}^T$, we have 60 eigentriples, which are ordered by their contribution (shares) in the decompositions, as well as 60 eigenvectors and principal components. With the window length $L = 60$ and use of the first 12 eigenvalues for reconstructing the original series without noise we consider remaining eigentriples (13-60) as noise.

We used SSA predictions, but the results of these predictions are less accurate than the predictions by MSSA. Therefore, we only report the results based on MSSA method. We used the following series in MSSA: CPI-all, CPI-core, CPI for Food and beverages, and CPI for Housing.

Table 4.19 shows the RMSEs for MSSA/random walk for 1-step and 3-step ahead forecasts of inflation rate based on the CPI-all and CPI-core series for a number of periods. The results indicate that MSSA outperforms the random walk predictions in both one and 3-step ahead

forecasts and in all time periods considered in the table.

Table 4.19 also presents the results of Diebold and Mariano test indicating whether the discrepancies between MSSA and RW model forecasting procedures are statistically significant. The results of this table confirm that, for all cases, the differences are significant at 1% confidence level.

Additionally, Table 4.19 presents test results for the null hypothesis of whether the percentages of the direction of changes are greater than the pure chance (50%). The table shows that all results are statistically significant at 1% confidence level. It should be noted that the MSSA prediction results for CPI-core series are better than MSSA prediction results for the CPI-all series. This maybe due to the higher volatility of prices of foods and energy which are excluded in construction of CPI-Core. The results of this table also show that MSSA predicts direction of change for 3-step as accurately as it can predict 1-step ahead.

Year	N	RMSE (SSA_h/RW_h)				Direction of Change			
		h=1		h=3		1-step ahead		3-step ahead	
		CPI-all	CPI-core	CPI-all	CPI-core	CPI-all	CPI-core	CPI-all	CPI-core
Jan 97–Dec 98	24	0.62*	0.52**	0.44**	0.34**	0.79**	0.79**	0.79**	0.79**
Jan 97–Dec 00	48	0.72*	0.55**	0.65**	0.35**	0.70**	0.77**	0.70**	0.79**
Jan 97–Dec 02	72	0.72**	0.51**	0.60**	0.35**	0.68**	0.80**	0.68**	0.83**
Jan 97–Dec 04	96	0.72**	0.51**	0.55**	0.35**	0.65**	0.80**	0.66**	0.82**
Jan 97–Dec 06	120	0.76**	0.50**	0.56**	0.33**	0.64**	0.80**	0.64**	0.81**
Jan 97–Nov 08	143	0.78**	0.53**	0.58**	0.35**	0.67**	0.79**	0.65**	0.82**

Table 4.19. RMSE of MSSA forecast results with respect to the RW method, Diebold-Mariano significance test results and direction of change test for inflation rate based on the CPI-all and CPI-core series. ** and * imply significance at 1% and 10% confidence levels, respectively.

To acquire a better understanding of forecasting accuracy of the methods, we examine the empirical cumulative distribution function for the absolute errors of the MSSA and RW methods next. Fig. 4.25 presents the empirical cumulative distribution function (CDF) for the absolute errors of the MSSA and RW forecasts. The graph on the left is for 1-step ahead and to the right is 3-step ahead predictions.

Fig. 4.25 shows that the empirical distribution of the RW errors stochastically dominates the distribution of the MSSA errors (that is, the RW errors are stochastically larger than the MSSA errors). It appears from Fig. 4.25 that the frequencies of larger errors for the random walk model are substantially higher compared to MSSA's errors. In fact, the maximum error for MSSA in both 1-step and 3-step prediction is 0.003, while the maximum error for the random walk is almost 0.008 for 3-step ahead prediction, and the maximum error for 1-step ahead is approximately 0.006.

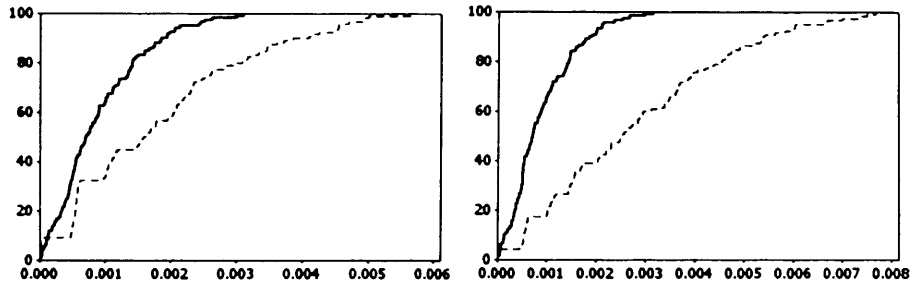


Figure 4.25. Empirical cumulative distribution functions of the absolute errors for MSSA (thick line) and random walk (dashed line) for 1-step ahead (left side) and 3-step ahead forecast (right side) over the period JAN 1997 to Nov 2008.

4.5.4 Comparison with the other methods

Comparative study is somewhat difficult, since data, methods, forecasting horizons and error criteria are not uniform. Nevertheless, we

compare the results based on the MSSA method with the results obtained from other inflation prediction methods. We smoothed the series by taking 3-month moving averages to make our results are comparable with prediction results of other inflation researchers who had smoothed the CPI data by the moving average method [92]. Table 4.20 presents RMSEs of MSSA prediction results and the forecasting results obtained using the other models considered in this study for 3-step ahead predictions [8].

[8]The labor intensive work on predicting 12-month and 24-month ahead predictions are underway, and we hope to present those results in future in another paper.

Year	N	MSSA/AO		MSSA/AR		MSSA/DFM88		MSSA/DFM158	
		CPI-all	CPI-core	CPI-all	CPI-core	CPI-all	CPI-core	CPI-all	CPI-core
Jan 97–Dec 98	24	0.54**	0.34**	0.57**	0.27**	0.48**	0.68**	0.58**	0.26**
Jan 97–Dec 00	48	0.73**	0.37**	0.79**	0.31**	0.75**	0.36**	0.78**	0.35**
Jan 97–Dec 02	72	0.68**	0.45**	0.73**	0.39**	0.75**	0.38**	0.77**	0.34**
Jan 97–Dec 04	96	0.70**	0.42**	0.70**	0.40**	0.73**	0.40**	0.74**	0.38**
Jan 97–Aug 06	116	0.75**	0.39**	0.76**	0.36**	0.79**	0.37**	0.81*	0.36**

Table 4.20. RMSE of MSSA forecast with other models for 3-month ahead forecast for 3-month moving averages of inflation rate based on the CPI-all and CPI-core series. AO= Atkeson and Ohanian; AR=Autoregressive; DFM88 =Dynamic factor model based on 88 variable; DFM158=Dynamic factor model based on 158 variables.

In Table 4.21, we present results for the direction of change in the moving average series according to all inflation forecasting methods discussed in this study. The numbers in the data show the percentage of time a method correctly predicted the direction of change in a series. The numbers indicate that MSSA method correctly predicts the direction of the change of the moving average of CPI-Core consistently higher than the competing models. This is particularly true for longer prediction horizons. For example, compare number 0.84 under MSSA-CPI-Core column for period of January 1997–August 2006, with the remaining entries in the same row. The superior performance of MSSA for CPI-all for period equal to or longer than 96 observations is apparent also. The statistical significance of the predicted values are also presented in the table.

Year	N	MSSA		AO		AR		DFM88		DFM158	
		CPI-all	CPI-core	CPI-all	CPI-core	CPI-all	CPI-core	CPI-all	CPI-core	CPI-all	CPI-core
Jan 97–Dec 98	24	0.87**	0.96**	0.89**	0.89**	0.91**	0.92**	0.92**	0.92**	0.89**	0.92**
Jan 97–Dec 00	48	0.73**	0.96**	0.73**	0.76**	0.80**	0.84**	0.84**	0.84**	0.80**	0.84**
Jan 97–Dec 02	72	0.70**	0.86**	0.57*	0.62**	0.72**	0.76**	0.75**	0.76**	0.69**	0.76**
Jan 97–Dec 04	96	0.71**	0.88**	0.43	0.50	0.56	0.61**	0.65**	0.61**	0.55	0.63**
Jan 97–AUG 06	116	0.72**	0.84**	0.28	0.37	0.44	0.53	0.56	0.52	0.44	0.58*

Table 4.21. Direction of change results of 3-month ahead forecasts of the moving average series. * and ** indicate the 10% and 1% levels of significance, respectively.

4.5.5 Inflation rate based on the GNP and GDP price index: 1970s to mid-1980s and 1985-2007

We compared the predictive accuracy of the SSA in prediction of inflation in two periods of 1970-1985 and 1985-2007 in the United States. The results based on GDP and GNP data appear in table 4.22 and indicate that SSA has a superior performance during the more stable inflationary period. We have better movement forecast (DC test) test results for 1985-present also. These results are expected since a RW model's predictive capacity diminishes for a high volatility time series.

RMSE(SSA/RW): GNP Deflator		RMSE(SSA/RW): GDP Deflator	
1970–1985	0.92**	1970–1985	0.91**
1985–2007	0.73**	1985–2007	0.71**
DC test: GNP Deflator		DC test: GDP Deflator	
1970–1985	0.62*	1970–1985	0.63*
1985–2007	0.81**	1985–2007	0.80**

Table 4.22. The results of DC test and the ratio of root mean squared error (RMSE) of SSA/random walk for the quarterly and annual real-time GNP and GDP chain-weighted price indexes. * and ** indicate the 10% and 1% levels of significance, respectively.

4.5.6 Discussions

We believe the present work has made two important contributions to the literature on forecasting inflation in the United States. First, we have documented, at least for the data used in this study, that MSSA method is a superior forecasting technique compared to the other methods used in the literature. Second, the predictive power of the MSSA vis-a-vis a RW model tends to diminish during the more volatile inflationary period. Our results are consistent with the findings of other researchers in predicting inflationary pressure in 1970-1985 and during

the " Great Moderation" periods of 1985-2008.

Many researchers use the inflation rate forecast produced by the staff of Research Department at the Board of Governors of the Federal Reserve System as a benchmark series in comparative analysis of inflation prediction. The forecasts of a large number of macroeconomics and financial variables appear in the Fed's Greenbook (GB). The Greenbook forecasts of price inflation include the quarterly forecasts for the GNP/GDP Price Index for 1965.4 to 2001.4 converted to annualized percentage rates and are available at the Federal Reserve Bank of Philadelphia's web site. The link to the site appears under Federal Reserve Bank of Philadelphia in the Reference section below.

Atkeson and Ohanian [88] find that RMSE=1.01 for Green Book predictions and their naive method's predictions (next year inflation equals the last year inflation), implying that on average the (GB) predictions have not been better than the naive method's, that is, a RW method predictions. Based on this observation, we do not directly compare SSA-based forecasts with the GB forecasts. However, by implication, one may conclude that MSSA forecasts would outperform the GB forecasts since according to the results of the present study, MSSA outperforms the RW forecasts and the GB forecasts do not outperform the forecasts of the U.S. inflation rates based on a RW model, at least up to 1996, the latest data used in [88].

4.6 Summary and Conclusion

The empirical results of this chapter indicate that SSA can be successfully applied to the analysis and forecasting of economic time series. This chapter has illustrated that the SSA technique performs well in

the simultaneous extraction of harmonics and trend components. The comparison of forecasting results showed that SSA is more accurate than several well-known methods, in the analysis and future prediction of different time series. The series considered here are some examples of a seemingly complex series with potential structure which can be easily analysed by SSA and could provide a typical example of a successful application of SSA.

The SSA, ARIMA and Holt-Winter methods were compared for forecasting seasonally unadjusted monthly data on industrial production indicators in Germany, France and the UK and also U.S accidental death series. We have demonstrated that SSA is a very powerful tool for analyzing and predicting these series. For industrial production series, the SSA technique outperformed the ARIMA and Holt-Winter methods in predicting the values of the production series according to the RMSE criterion, particularly at horizons of $h = 3, 6$ and 12 months. We have also found that SSA works well for small sample sizes, as for the UK with the sample size of 84 observations. The forecasts obtained by bootstrapping also confirm the findings. As the results show, the three methods perform similarly well in predicting the direction of change. However, SSA outperforms the Holt-Winter and ARIMA models at longer horizons and hence can be considered as a reliable method for predicting recessions and expansions.

To examine capability of SSA in forecasting short time series. We used 32 Iranian national account data sets describing the main economic features of the Islamic Republic of Iran, as provided on the website of the Central Bank of the Islamic Republic of Iran. The data are given in a quarterly and yearly format and have different types of

non-stationarity. All the data sets are rather short. The results show that SSA can be successfully used for the analysis and forecasting of short economic time series with different types of non-stationarity. In particular, many quarterly series have periodic components with non-stationary amplitudes but SSA has been able to extract and forecast these periodic components very accurately. Most of the yearly data have clear structural changes which makes the application of standard methods of analysis almost impossible.

Another important finding is that unlike standard methods used for analysis of economics time series, SSA does not require parametric models or transformation of the data into the logarithmic scale. Moreover, our study has shown that in most cases, the transformation of the quarterly series into logarithmic scale has lead to the deterioration of the precision of the forecasts.

To evaluate multivariate varsion of SSA, MSSA, the univariate and multivariate SSA were used in prediction of value and direction of changes (series moving up or down) in the daily UK exchange rates. A random walk model was utilized as a benchmark model to compare performances of the SSA, MSSA, and direction of change criterion in these prediction exercises. We employed Diebold-Mariano and the direction of change criteria to validate the findings. The empirical results and the test statistics show that MSSA have outperformed random walk models for the pound / dollar exchange rate series (similar results were obtained for the euro/dollar series, but we do not report them in this paper). As was pointed out earlier, UK and EU move in proximity of each other and have high correlation. The high correlation between two series is a good indicator of accurate predictability of one series using

the two series together in prediction exercises.

The use of traditional time series analysis of unit root test confirm that both UK and EU series are non-stationary series. We tested for cointgeration and found that the series are cointegrated. We further estimated an error correction (EC) model for the cointgerated series and used it for prediction. The prediction results based on EC model show an inferior performance compared to predictions by a RW as well as SSA and MSSA methods. The Granger causality test confirms that there exists a two-way causality between pound/dollar and EU/dollar exchange rates.

Given that the traditional structural econometric models of exchange rates have a poor record in prediction of the exchange rates in comparison to random walk models, we believe SSA and MSSA methods are highly promising. As is shown in this paper, the SSA method, at least in its multivariate representation, has decisively outperformed random walk models for exchange rate series. Further methodological development in this field as well as extensive application of these methods in financial and economic data could prove to be indispensable for accurate prediction exercises.

Several price indexes including consumer price index with and without highly volatile food and energy items as well as quarterly Chain-weighted GDP and GNP price indexes were also used for forecasting inflation rate and price levels.

The results of this research indicate that the SSA significantly outperforms all other methods commonly used in inflation forecasting. The superior prediction results are based on the capability of the SSA method to discard the stochastic components of the original series.

The results show that without exception, SSA outperforms both the naive random walk method and more complex econometric models that are used by other researchers in forecasting inflation rate based on the GDP price index. Moreover, we find that MSSA outperforms the random walk predictions in both one and 3-step ahead forecasts as well as all other time periods considered for forecasting inflation rate based on the CPI-all and CPI-core series (see Table 4.19). The Diebold and Mariano tests also confirm that, for all cases, the results are significant at 1% confidence level. We also find that SSA performs very well in predicting the direction of change. Additionally, we find that the empirical distribution of the RW errors also stochastically dominates the distribution of the MSSA errors for one and 3-step ahead forecast (see Fig. 4.25). Diebold and Mariano tests also confirm that the discrepancy between MSSA and RW model forecasting procedures are statistically significant at 1% confidence level.

The MSSA forecasting results were also compared with those results obtained by Phillips curve, DFM and AR(p) models. Once again, MSSA outperforms all other models for forecasting inflation rate and direction of change in the CPI-all and CPI-core (see Tables 4.20 and 4.21).

Finally, in light of inadequate performances of the NAIRU Philips curve-based and the time series models, we conclude that using SSA and MSSA is more promising for obtaining accurate forecasting of inflation rate.

Chapter 5

SSA BASED ON THE MINIMUM VARIANCE ESTIMATOR

In this chapter, the SSA technique based on the minimum variance estimator is introduced. The SSA technique based on the minimum variance and least squares estimators in reconstructing and forecasting time series are also considered. The monthly accidental deaths in the USA time series is used in examining the performance of the technique. The results are compared with several classical methods namely, Box-Jenkins SARIMA models, the ARAR algorithm and the Holt-Winter algorithm.

5.1 Introduction

The results of previous chapter confirm that errors can seriously limit the performance of the time series analysis methods and techniques. The previous works also indicate that the SVD based methods and signal subspace (SS) methods are very effective in filtering and forecasting the financial and economics time series [5].

Having a method for decomposing the vector space of the noisy time series into a subspace that is generated by the noise free series and a subspace for the noise component, we can construct the noise free time series. Approximate decomposition of the vector space of the noisy time series into noise free time series and noise series subspace can be done with, for example, the orthogonal matrix factorization technique such as SVD.

The idea to perform SS method was proposed in [99] where a modified SVD is used for reconstruction of noise free series. A general framework for recovering noise free series has been presented in [100]. The method forms the basis for a very general class of subspace-based noise reduction algorithms, is based on the assumption that the original time series exhibits some well-defined properties or obeys a certain model. Noise free series is therefore obtained by mapping the original time series onto the space of series that possess the same structure as the noise free series.

In this context, the SSA technique which is SVD and SS based method, can be considered as a proper method for noise reduction and forecasting time series data sets.

The SSA algorithms that has been considered in literature are based on the standard SVD and the least squares (LS) estimate (see, for example, [8] and references therein). The LS estimate of the signal component is obtained by truncating the singular values of the noisy series. The LS estimator projects the noisy time series onto the perturbed signal (noise + signal) subspace. The reconstructed series using LS estimator has the lowest possible (zero) signal distortion and the highest possible residual noise level. In this chapter, we consider an alternative

method which is based on the minimum variance (MV) estimator for reconstruction and forecasting noisy time series. The MV estimator is the optimal linear estimator, which gives the minimum total residual power [101,102].

5.2 LS and MV Estimators

Consider a noisy vector Y_T of length T. To construct the noisy vector Y_T we will add the additive white noise N_T to the signal component S_T and assume that the noise is uncorrelated with the signal:

$$Y_T = S_T + N_T; \tag{5.2.1}$$

Define the so-called 'trajectory matrix' $\mathbf{X} = (x_{ij})_{i,j=1}^{L,K}$, where $x_{ij} = y_{i+j-1}$. It is obvious that:

$$\mathbf{X} = \mathbf{S} + \mathbf{N}, \tag{5.2.2}$$

where $\mathbf{S}$ and $\mathbf{N}$ represent Hankel matrices of the signal S_T and noise N_T, respectively. The SVD of the trajectory matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{5.2.3}$$

where $\mathbf{U} \in \mathbb{R}^{L\times K}$ is the matrix consists of the normalized eigenvector U_i corresponding to the eigenvalue λ_i $(i = 1, \ldots, L)$, $\mathbf{V} \in \mathbb{R}^{K\times K}$, is the matrix contains the principal components defined as $V_i = \mathbf{X}^T U_i/\sqrt{\lambda_i}$, and $\mathbf{\Sigma} = \mathrm{diag}(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L)$.

The SS methods are based on the assumption that the vector space of the noisy time series (signal) can be split in mutually orthogonal noise

and signal+noise subspaces. The components in the noise subspace are suppressed or even removed completely. Therefore, one can reconstruct the noise free series from signal+noise subspace by choosing the weight. Thus, by adapting the weights of the different singular components, an estimate of the Hankel matrix $\mathbf{X}$, which corresponds to noise reduced series, can be achieved:

$$\mathbf{X} = \mathbf{U}(\mathbf{W}\boldsymbol{\Sigma})\mathbf{V}^T, \tag{5.2.4}$$

where $\mathbf{W}$ is the diagonal matrix containing the weights. Now, the problem is choosing the weight matrix $\mathbf{W}$. Next we consider the problem of choosing this matrix using different criteria. The SVD of the matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = [\mathbf{U}_1 \;\; \mathbf{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 \\ 0 & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} \tag{5.2.5}$$

where $\mathbf{U}_1 \in \mathbb{R}^{L\times r}$, $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{r\times r}$ and $\mathbf{V}_1 \in \mathbb{R}^{K\times r}$. We can also represent SVD of the Hankel matrix of the signal $\mathbf{s}_T$ as:

$$\mathbf{S} = [\mathbf{U}_{1s} \;\; \mathbf{U}_{2s}] \begin{bmatrix} \boldsymbol{\Sigma}_{1s} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_{1s}^T \\ \mathbf{V}_{2s}^T \end{bmatrix} \tag{5.2.6}$$

It is clear that the Hankel matrix $\mathbf{S}$ can not be reconstructed exactly if it is perturbed by noise.

5.2.1 LS Estimate of $\mathbf{S}$

Let us consider the assumption that the matrix $\mathbf{X}_{L\times K}$ is rank deficient, i.e., rank $\mathbf{X} = r$ and $r < L < K$. The simplest estimate of $\mathbf{S}$ is obtained

when we approximate $\mathbf{S}$ by a matrix of rank r in the LS sense:

$$\min \| \mathbf{X} - \hat{\mathbf{S}}_{LS} \|_F^2 \tag{5.2.7}$$

where $\| \cdot \|_F$ is the *Frobenius* norm. That is, the LS estimate is obtained by setting the smallest singular value to zero ($\lambda_{r+1} = 0, \ldots, \lambda_L = 0$) in (6.2.7):

$$\hat{\mathbf{S}}_{LS} = [\mathbf{U}_1 \;\; \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T \tag{5.2.8}$$

The $\mathbf{S}_{LS}$ estimate removes the noise subspace, but keeps the noisy signal uncorrelated in the signal+noise subspace. Among different weighting methods, the LS estimate contains the highest possible residual noise level, only the noise from the noise subspace is filtered out, but has the lowest signal distortion (it keeps signal+noise subspace). The disadvantage of LS is that the performance of the LS estimator is crucially dependent on the estimation of the signal rank r. That is, selecting singular values in LS is a binary approach. The main advantage of the LS estimate is that one does not need to consider any assumptions either about the signal or noise. For example, if the noise is not white, many other methods need prewhitening and dewhitening steps [103].

5.2.2 MV Estimate of S

The aims of the noise reduction can be considered as follows: (1) separate the signal+noise subspaces from the noise only subspace, (2) remove the noise subspace, (3) ideally, remove the noise components in

the signal + noise subspace. The first two steps can be achieved by the least squares estimate, while the MV estimate allows us to have the third one as well. However, one should consider the following assumptions to obtain the MV estimate:

i) The signal is orthogonal to the noise: $\mathbf{S}^T\mathbf{N} = \mathbf{0}$.

ii) $\mathbf{N}^T\mathbf{N} = \sigma^2_{noise}\mathbf{I}$, where $\mathbf{I}$ is a identity matrix. That is, every column of $\mathbf{N}$ has norm σ_{noise}.

iii) The smallest singular value of $\mathbf{\Sigma}_1$, λ_r, is larger than largest singular value of $\mathbf{\Sigma}_2$, λ_{r+1}, where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are introduced in 5.2.5 and 5.2.6.

If the assumptions i–iii are met, one can obtain the MV estimate as described in [101, 102]. Let us consider the matrix $\mathbf{X}$, with rank $(\mathbf{X})$=rank $(\mathbf{N}) = L$ and also rank $(\mathbf{S}) = r$. Find the matrix $\mathbf{T} \in \mathbb{R}^{K\times K}$ that minimizes:

$$\min \| \mathbf{XT} - \mathbf{S} \|_F^2 . \tag{5.2.9}$$

The solution is obtained by

$$\mathbf{T} = (\mathbf{X^T X})^{-1}\mathbf{X^T S}. \tag{5.2.10}$$

Therefore, the MV estimate of $\mathbf{S}$ is:

$$\mathbf{XT} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T S}. \tag{5.2.11}$$

Using the SVD of the $\mathbf{X}$, we can obtain:

$$\mathbf{XT} = \mathbf{U}\mathbf{U}^T\mathbf{S}. \tag{5.2.12}$$

That is, the MV estimate of $\mathbf{S}$ can be interpreted as a orthogonal pro-

jection of $\mathbf{S}$ onto the column space of $\mathbf{X}$ because $\mathbf{U}\mathbf{U}^T$ is the associated projection matrix. Note also that rank $(\mathbf{XT})$= rank $(\mathbf{S}) = r$. In real application the matrix $\mathbf{S}$ is not known, but it is possible to achieve the MV estimate, from SVD of $\mathbf{X}$, if assumption i–iii are satisfied. Let us now consider an alternative form of the SVD of the matrix $\mathbf{X}$ using the SVD of $\mathbf{S}$ (5.2.6) as follows:

$$\begin{aligned}\mathbf{X} &= \mathbf{S} + \mathbf{N} = \mathbf{U}_{1s}\boldsymbol{\Sigma}_{1s}\mathbf{V}_{1s}^T + \mathbf{N}\mathbf{V}_{1s}\mathbf{V}_{1s}^T + \mathbf{N}\mathbf{V}_{2s}\mathbf{V}_{2s}^T \\ &= \left[(\mathbf{U}_{1s}\boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s})(\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2\mathbf{I})^{-1/2} \quad \sigma_{noise}^{-1}\mathbf{N}\mathbf{V}_{2s}\right] \\ &\times \begin{bmatrix} (\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2\mathbf{I})^{1/2} & \mathbf{0} \\ \mathbf{0} & \sigma_{noise}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{1s}^T \\ \mathbf{V}_{2s}^T \end{bmatrix}.\end{aligned} \tag{5.2.13}$$

As it appears from (5.2.13), the middle matrix is diagonal, and the left and right matrices have orthonormal columns. Therefore, (5.2.13) can be considered as an alternative form of the SVD of $\mathbf{X}$, and the singular values of $\mathbf{X}$ are:

$$\begin{aligned}\boldsymbol{\Sigma}_1 &= (\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2\mathbf{I})^{1/2}, \\ \boldsymbol{\Sigma}_2 &= \sigma_{noise}\mathbf{I}.\end{aligned} \tag{5.2.14}$$

Hence, the singular values in $\boldsymbol{\Sigma}_2$ can be considered for identification of noise threshold, which permits estimating σ_{noise} from $\boldsymbol{\Sigma}_2$ in (5.2.14). We can also consider the following submatrices:

$$\begin{aligned}\mathbf{U}_1 &= (\mathbf{U}_{1s}\boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s})(\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2\mathbf{I})^{-1/2} \\ &= (\mathbf{U}_{1s}\boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s})\boldsymbol{\Sigma}_1^{-1}, \\ \mathbf{U}_2 &= \sigma_{noise}^{-1}\mathbf{N}\mathbf{V}_{2s}, \\ \mathbf{V}_1 &= \mathbf{V}_{1s}, \\ \mathbf{V}_2 &= \mathbf{V}_{2s}.\end{aligned} \tag{5.2.15}$$

Now, using (5.2.14–5.2.15) and also $\mathbf{S}^T\mathbf{N} = 0$, $\mathbf{U}_1^T\mathbf{N} = 0$, we then

obtain the MV estimate of $\mathbf{S}$:

$$\begin{aligned}
\hat{\mathbf{S}}_{MV} &= \mathbf{U}\mathbf{U}^T\mathbf{S} = \mathbf{U}_1\mathbf{U}_1^T\mathbf{U}_{1s}\mathbf{\Sigma}_{1s}\mathbf{V}_{1s}^T + \mathbf{U}_2\mathbf{U}_2^T\mathbf{U}_{1s}\mathbf{\Sigma}_{1s}\mathbf{V}_{1s}^T \\
&= \mathbf{U}_1\mathbf{\Sigma}_1^{-1}(\mathbf{\Sigma}_{1s}\mathbf{U}_{1s}^T + \mathbf{V}_{1s}^T\mathbf{N}^T)\mathbf{U}_{1s}\mathbf{\Sigma}_{1s}\mathbf{V}_{1s}^T \\
&+\sigma_{noise}^{-1}\mathbf{U}_2\mathbf{V}_{2s}^T\mathbf{N}^T\mathbf{U}_{1s}\mathbf{\Sigma}_{1s}\mathbf{V}_{1s}^T \qquad (5.2.16) \\
&= \mathbf{U}_1\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{1s}^2\mathbf{V}_{1s}^T \\
&= \mathbf{U}_1\mathbf{\Sigma}_1^{-1}(\mathbf{\Sigma}_1^2 - \sigma_{noise}^2\mathbf{I})\mathbf{V}_1^T .
\end{aligned}$$

5.2.3 Weight matrix $\mathbf{W}$

Let us consider again the weight matrix $\mathbf{W}$ based on the LS and MV estimates. As it is appears from (5.2.8) and (5.2.16), the left and right singular vector, $\mathbf{U}_1$ and $\mathbf{V}_1$, of LS and MV estimates are the same, but the singular values are different. The LS and MV estimates can be defined based on the weight matrix $\mathbf{W}_{r\times r}$ as follows:

$$\begin{aligned}
\hat{\mathbf{S}}_{LS} &= \mathbf{U}_1(\mathbf{W}_{LS}\mathbf{\Sigma}_1)\mathbf{V}_1^T \\
\hat{\mathbf{S}}_{MV} &= \mathbf{U}_1(\mathbf{W}_{MV}\mathbf{\Sigma}_1)\mathbf{V}_1^T
\end{aligned} \qquad (5.2.17)$$

where

$$\begin{aligned}
\mathbf{W}_{LS} &= \mathbf{I}_{r\times r} \\
\mathbf{W}_{MV} &= \mathrm{diag}\left((1 - \tfrac{\sigma_{noise}^2}{\lambda_1^2}), \ldots, (1 - \tfrac{\sigma_{noise}^2}{\lambda_r^2})\right)
\end{aligned} \qquad (5.2.18)$$

5.3 Separability

The success of the SSA technique based on the MV estimate, essentially depends on the assumptions i–iii, which in practice, except probably for condition iii, are never satisfied exactly. Let us consider the first assumption. If, for example, $\mathbf{S}^T\mathbf{N} \neq 0$ but $||\mathbf{S}^T\mathbf{N}||$ is small, we can still use the SVD of $\mathbf{X}$. The smaller $\mathbf{S}^T\mathbf{N}$ gets, the better will be

the approximations. For the second assumption we can assume that $E(\mathbf{N}^T\mathbf{N}) = \sigma^2_{noise}\mathbf{I}$. However, it has been shown that, under some weak conditions, the assumptions i–iii can be considered true asymptotically [101]. We can therefore still use the robustness feature of SVD with respect to weak violations of these conditions. To overcome this problem we use the concept of separability, that can be considered instead of the above conditions (see chapter 2 and also [8]).

If the absolute value of the **w**-correlation is small, then the two series are almost **w**-orthogonal, but, if it is large, then the series are far from being **w**-orthogonal and therefore we have weak separability (for more information see chapter 4).

We shall say that the series S_T and N_T are *asymptotically separable* if the maximum $\rho^{(L,K)}$ of the absolute values of the correlation coefficients between the rows/columns of the trajectory matrices of the series S_T and N_T tends to zero, as $T \to \infty$ (for further information see Appendix F).

From the practical viewpoint, the effect of the asymptotic separability becomes apparent in the analysis of long series and means that two asymptotically separable series are approximately separable for large T. For several analytical examples of both exact and asymptotic separability see [8]. It should be noted that the class of asymptotically separable series is much wider than the class of series that are exactly separable, and the conditions on the choice of the window length L are much weaker in the case of asymptotic separability.

Conditions for asymptotic separability are much weaker. In particular, two harmonics with arbitrary different frequencies are asymptotically separable as soon as L and K tend to infinity. Moreover, under the

same conditions the periodic components are asymptotically separable from the trends of a general form.

5.4 Empirical results and comparison

5.4.1 Simulated series

We shall consider two types of time series; real and artificially generated time series. The capability of the SSA technique based on the minimum variance estimator, in reconstructing and forecasting, was initially tested by applying the technique to the simple sin series:

$$y_t = Sin(2t\pi/12) + \epsilon_t \tag{5.4.1}$$

where ϵ_t is a white noise series. In total 300 data are generated and we added different normally distributed noise to each point of the original series. The simulation was repeated 1000 times. The first 200 observations was considered as in-sample (reconstruction) and the rest as out-of-sample series (forecasting). We also considered different values of window length L to examine the sensitivity of the SSA technique for different L.

Let us first consider **w**-*correlation* between reconstructed signal and noise series. Figure 5.1 shows **w**-*correlation* between extracted signal using SSA based on the minimum variance (SSA_{MV}) and least squares (SSA_{LS}). As the figure shows, **w**-*correlation* tends to zero as the window length increases confirming theoretical results mentioned in previous section. We also used normality test to examine whether the noise components are distributed normally or not. The results of normality test confirms that all noise components are distributed normally.

Figure 5.2 shows the RMSE of reconstructed series (left) and also forecasted series (right). As appears from Figure 5.2, the SSA_{MV} has slightly better performance in both reconstruction and forecasting for small window length; the performance of both estimate are similar for a large window length.

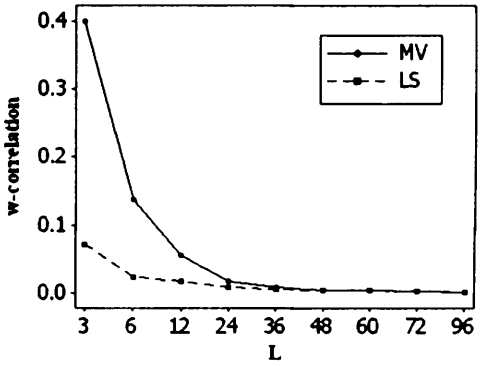


Figure 5.1. w-correlation between extracted signal and noise series for different window length based on the LS (dashed line) and MV estimate (thick line).

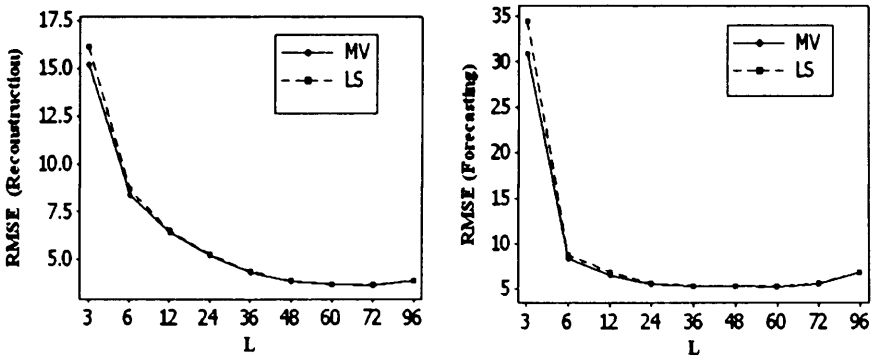


Figure 5.2. The performance of SSA_{MV} and SSA_{LS} in reconstruction (left) and forecasting (right) noisy sin series.

So far, we considered the situation where the noise component ϵ_t is distributed normally. Next we consider if ϵ_t is not distributed normally. To calculate the precision we use the ratio of RMSE (RRMSE).

Figure 5.3 shows the RRMSE of SSA_{MV}/SSA_{LS} in reconstruction

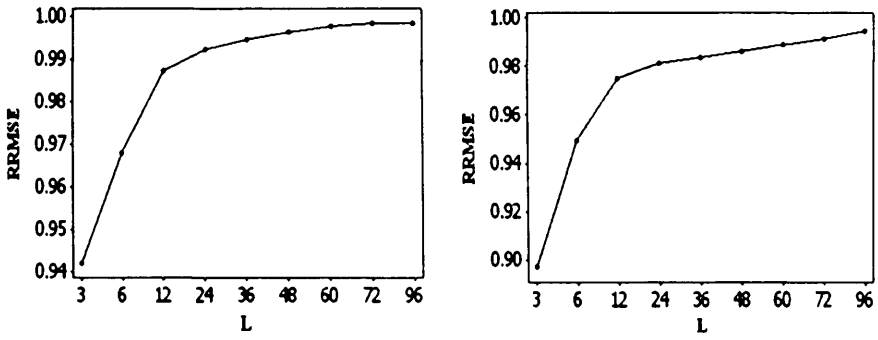


Figure 5.3. The RRMSE of $\text{SSA}_{MV}/\text{SSA}_{LS}$ in reconstruction (left) and forecasting (right) noisy sin series.

(left) and forecasting (right) noisy sin series, where the noise term is distributed non-normally. As the figure shows, RRMSE tends to 1 as the window length increases confirming that both methods have similar performance for large window length. The graphs also show that there is a gradual increase in RRMSE with window length. For window length 3, the performance of SSA_{MV} is up to 10% better than SSA_{LS} in forecasting, while this is approximately 6% in reconstruction. However, there is not a significant discrepancy between the performance of SSA_{MV} and SSA_{LS} for window length greater than 12.

5.4.2 Real series

Let us now consider the performance of the SSA technique based on the MV and LS estimates by applying it to a well-known time series data set, namely, monthly accidental deaths in the USA.

The window $L = 24$ and the first 12 singular values have been used in reconstructing and forecasting the series y_t and singular values 13–24 have been considered as noise components (for more information about parameters selection, for this series, see previous chapter and [14]). Here, we used the same parameters and recurrent forecasting algorithm

as for the vector forecasting algorithm that was used in [14].

The methods are arranged based on the performance of forecasting. The results are presented in Table 5.1. The values of RMSE show performance of forecasting. The last six columns, labeled RRMSE, show the ratios of RMSEs SSA/other competitive methods. As it appears in Table 5.1, the forecasting performance using the SSA technique based on the LS estimate (SSA_{LS}) and based on the MV estimate (SSA_{MV}) are much better than other forecasting methods and also the SSA_{MV} is the best among the methods considered, for example, the value of RMSE for the SSA_{MV} is 9 times less than the first one (model I) and almost 3 times less than the ARAR algorithm. From the table, one can see that the SSA_{MV} performance is better that the SSA_{LS}. Let us consider the performance of the SSA forecasting results with respect to different values of r. We choose the same window length L but different number of eigenvalues r. The results are presented in Table 5.1, for the first 13 and 14 eigenvalues. As the table shows, again, the SSA technique outperforms the other classical methods.

Table 5.1. RRMSE of the post-sample forecasts.

Method	RMSE	RRMSE					
		r=12		r=13		r=14	
		SSA_{LS}	SSA_{MV}	SSA_{LS}	SSA_{MV}	SSA_{LS}	SSA_{MV}
Model I	582.63	0.21	0.12	0.18	0.11	0.27	0.13
Model II	500.50	0.24	0.14	0.21	0.13	0.31	0.16
H-W	401.26	0.30	0.18	0.26	0.17	0.39	0.20
ARAR	253.20	0.47	0.28	0.41	0.26	0.62	0.31

It can be seen that the quality of the forecast is changed when one changes the number of eigenvalues in the reconstruction step. Of course, forecasting accuracy and reconstruction quality are related. By

selecting a group of eigenvalues, and considering other eigenvalues as noise, some frequencies may be filtered out completely. This destroys the signal structure and then gives a poorer reconstruction. In general, a high signal to noise ratio will result in accurate forecasting and vice-versa. Let us consider w-correlation between extracted signal and noise component and normality test of noise series for different group of eigentriples ($r = 12, 13$ and 14). Table 5.2 represents the results. As can be seen from the table, the signal and noise component separated very well as w-correlation is very small (0.006, 0.005 and 0.004). Therefore, the assumption of orthogonality can be accepted here and consequently we can use MV estimate. Here we used different normality test; Anderson-Darling (A-D), Ryan-Joiner (R-J) and Kolmogorov-Smirnov (K-S). The symbol ** indicates the results at the 1% level of significance. The normality test results indicate that the assumption of normality of noise component is acceptable (which is an essential criterion in MV estimate). It should be noted that we need the assumption of normality only for SSA based on the MV estimator and we do not need any assumptions for SSA based on the LS estimator.

Table 5.2. w-correlation and normality test.

		$r = 12$	$r = 13$	$r = 14$
w-correlation		0.006	0.005	0.004
	A-D	0.25**	0.19**	0.36**
Normality test	R-J	0.99**	0.99**	0.99**
	K-S	0.07**	0.05**	0.08**

5.5 Conclusion

Classical time series methods such ARIMA type models fit a model directly from noisy data and use the fitted model for forecasting future

data points. Forecasting results are typically better if one fits a model to noise reduced time series and then use the fitted model for forecasting new data points. The signal subspace and SVD based methods such as SSA can be applied as powerful tools for finding the noise free series and using it for forecasting future data points.

In this chapter we introduced the SSA technique based on the minimum variance estimator. The comparison of the forecasting results showed that SSA, based on the minimum variance (MV) and structured total least squares (LS) estimates, are much more accurate than several well-known classical methods, in forecasting of a well know time series. We also found that the SSA forecasting results based on MV are better than based on LS for considered series. However, comparison between these two estimates depends on the choice of the SSA parameters, the window length L and the number of eigenvalues r, the data we have and also the analysis we have to perform. In conclusion, the results confirm that the SSA technique based on both estimates, LS and MV, gives much more accurate results than the classical methods of time series analysis considered here.

Chapter 6

SSA BASED ON THE PERTURBATION THEORY

In this chapter, we consider the SSA technique based on the perturbation theory. The performance of the SSA technique based on the perturbation theory is assessed in reconstructing and forecasting different time series (stationary and non-stationary). The performance of the proposed algorithm is assessed with respect to different window length L and different values of the signal to noise ratio. For consistency with the results obtained in previous chapters, the USA death series, financial time series, and chaotic series are used to evaluate the performance of the proposed technique. The results are also compared with several classical methods.

6.1 Introduction

Consider a noisy signal vector $\widetilde{Y}_T$ of length T. Let us add the additive noise to the noise free series (signal) Y_T and assume that the noise series δY_T is uncorrelated with the signal:

$$\widetilde{Y}_T = Y_T + \delta Y_T; \tag{6.1.1}$$

Note that we use different notation in this chapter to emphasis that the perturbed term is not always noise. For example, the perturbed term can be a harmonic component with different amplitude added to the original series. This is the general idea behind on the perturbation theorem. Moreover, we will show latter that the structure of the trajectory matrix is different with those represented in previous chapter. Therefore, we kept this notation. Let us define trajectory matrix $\widetilde{\mathbf{X}} = (\widetilde{x}_{ij})_{i,j=1}^{L,K}$, where $\widetilde{x}_{ij} = \widetilde{y}_{i+j-1}$. It is clear that:

$$\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}, \tag{6.1.2}$$

where $\mathbf{X}$ and $\delta\mathbf{X}$ represent Hankel matrices of the signal Y_T and noise δY_T, respectively.

As we mentioned in previous chapter, the LS estimator projects the noisy time series onto the perturbed signal (noise + signal) subspace. Therefore, the reconstructed series still have some part of the initial noise level δX_N due to the nature of LS estimate. In this chapter we introduce another alternative technique to overcome this problem. In the previous chapter we used SSA based on the minimum variance estimator which produced a better approximation of matrix $\mathbf{\Sigma}$. As the results showed, this improvement helps us to have a better reconstruction. However, the left and right eigenvectors (U_i and V_i) are still noisy and have some part of the noise component. In this chapter, we try to overcome this problem by means of perturbation theory. That is, we represent a better approximation of matrices $\mathbf{\Sigma}$, $\mathbf{U}$ and $\mathbf{V}$. This will help us to reconstruct the signal matrix better as we remove those parts of noise components from $\widetilde{\mathbf{\Sigma}}$, $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$. Therefore, we expect the SSA

based on LS estimate gives the lowest forecasting accuracy and based on the perturbation theory yields the highest performance.

The next section briefly describes perturbation theory and its application for subspace methods. The improvement of the SSA technique based on the perturbation theory is considered in section 3. The empirical results are then presented and described in Section 4 and some conclusions are given in Section 5.

6.2 Perturbation Theory

6.2.1 Related theorems

Let us now consider the problem of separation of an additive noise component from a perturbation theory point of view. First we consider some useful theorems.

Theorems 1

Let $\mathbf{X}$ and $\delta\mathbf{X}$ be Hermitian matrices and $\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$. Let the eigenvalues of $\mathbf{X}$ be $\lambda_1 \geqslant \ldots \geqslant \lambda_L$, and let the eigenvalues of $\widetilde{\mathbf{X}}$ be $\widetilde{\lambda}_1 \geqslant \ldots \geqslant \widetilde{\lambda}_L$. If μ_L is the smallest eigenvalue of $\delta\mathbf{X}$, then [104]

$$\widetilde{\lambda}_i \geqslant \lambda_i + \mu_L \qquad i = 1, \ldots, L \tag{6.2.1}$$

There are two useful characteristics about the above theorem; it restricts the location of the eigenvalus of the perturbed matrix $\widetilde{\mathbf{X}}$, but there is no restriction on the size of the perturbation $\delta\mathbf{X}$. Some perturbation bounds of the singular values has been considered in [105] as follows.

Theorems 2

Perturbation bounds for the singular values of $L \times K$ matrix $\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ are [105]

$$|\widetilde{\lambda}_i - \lambda_i| \leq ||\delta\mathbf{X}||_2$$
$$\sum_{i=1}^{L}(\widetilde{\lambda}_i - \lambda_i) \leq ||\delta\mathbf{X}||_F^2 \tag{6.2.2}$$

where $||.||_2$ and $||.||_F$ are Euclidean norm and Frobenius norm, respectively. The above conditions indicate that the eigenvalues of the matrix $\mathbf{X}$ are well-conditioned with respect to perturbations. That is, perturbations of y_t produce similar or smaller perturbations in the singular values [105].

6.2.2 Subspace method and perturbation theory

Consider the following matrix

$$\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X} \tag{6.2.3}$$

where $\widetilde{\mathbf{X}}$ is a perturbed version of $\mathbf{X}$ with perturbation $\delta\mathbf{X}$. The SVD of the matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = [\mathbf{U}_s \;\; \mathbf{U}_n] \begin{bmatrix} \mathbf{\Sigma}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^T \\ \mathbf{V}_n^T \end{bmatrix} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^T \tag{6.2.4}$$

where $\mathbf{U}_s \in \mathbb{R}^{L\times r}$, $\mathbf{\Sigma}_s \in \mathbb{R}^{r\times r}$ and $\mathbf{V}_s \in \mathbb{R}^{K\times r}$.

Note that in permutation theory, the structure of the above matrix is different. In general case, we use matrix $\mathbf{\Sigma}_n$ in place of zero matrix in (6.2.6). That is we have the following structure:

$$\mathbf{X} = [\mathbf{U}_s \;\; \mathbf{U}_n] \begin{bmatrix} \mathbf{\Sigma}_s & 0 \\ 0 & \mathbf{\Sigma}_n \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^T \\ \mathbf{V}_n^T \end{bmatrix} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^T + \mathbf{U}_n\mathbf{\Sigma}_n\mathbf{V}_n^T \quad (6.2.5)$$

But, here we consider this part of the matrix as null space (zero matrix), which is similar to basic SSA terminology. Thus, the SVD of the matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = [\mathbf{U}_s \;\; \mathbf{U}_n] \begin{bmatrix} \mathbf{\Sigma}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^T \\ \mathbf{V}_n^T \end{bmatrix} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^T \quad (6.2.6)$$

The matrices $\mathbf{U}_s$ and $\mathbf{V}_s$ span the column spaces of $\mathbf{X}$ and $\mathbf{X}^T$, respectively, whereas $\mathbf{U}_n$ and $\mathbf{V}_n$ span their orthogonal spaces. Similarly, the SVD of the matrix $\widetilde{\mathbf{X}}$ can be written as:

$$\widetilde{\mathbf{X}} = [\widetilde{\mathbf{U}}_s \;\; \widetilde{\mathbf{U}}_n] \begin{bmatrix} \widetilde{\mathbf{\Sigma}}_s & 0 \\ 0 & \widetilde{\mathbf{\Sigma}}_n \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{V}}_s^T \\ \widetilde{\mathbf{V}}_n^T \end{bmatrix} = \widetilde{\mathbf{U}}_s\widetilde{\mathbf{\Sigma}}_s\widetilde{\mathbf{V}}_s^T + \widetilde{\mathbf{U}}_n\widetilde{\mathbf{\Sigma}}_n\widetilde{\mathbf{V}}_n^T \quad (6.2.7)$$

It is clear that the SVD of the matrix $\widetilde{\mathbf{X}}$ is completely different from the SVD of the matrix $\mathbf{X}$ due to perturbation term $\delta\mathbf{X}$. Next the aim is to derive general expressions for approximations to the perturbed terms up to the second order of $\delta\mathbf{X}$. Assume the perturbed terms are as follows:

$$\begin{aligned} \widetilde{\mathbf{U}}_s &= \mathbf{U}_s + \delta\mathbf{U}_s = \mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1 + \mathbf{U}_s\mathbf{P}_2 \\ \widetilde{\mathbf{V}}_s &= \mathbf{V}_s + \delta\mathbf{V}_s = \mathbf{V}_s + \mathbf{V}_n\mathbf{P}_3 + \mathbf{V}_s\mathbf{P}_4 \end{aligned} \quad (6.2.8)$$

$$\begin{aligned}\widetilde{\mathbf{U}}_n &= \mathbf{U}_n + \delta\mathbf{U}_n = \mathbf{U}_n + \mathbf{U}_s\mathbf{Q}_1 + \mathbf{U}_n\mathbf{Q}_2 \\ \widetilde{\mathbf{V}}_n &= \mathbf{V}_n + \delta\mathbf{V}_n = \mathbf{V}_n + \mathbf{V}_s\mathbf{Q}_3 + \mathbf{V}_n\mathbf{Q}_4\end{aligned} \tag{6.2.9}$$

$$\widetilde{\mathbf{\Sigma}}_s = \mathbf{\Sigma}_s + \delta\mathbf{\Sigma}_s \tag{6.2.10}$$

Note that the perturbed terms in (6.2.8) and (6.2.9) consists of two parts; the first part captures the perturbation in its orthogonal space and the second part considers perturbation in each subspace. Let us, for example, consider the perturbation term $\widetilde{\mathbf{U}}_s$. The perturbation term $\delta\mathbf{U}_s$ consists of two parts; $\mathbf{U}_n\mathbf{P}_1$ which captures the perturbation in its orthogonal space $\mathbf{U}_n$ and $\mathbf{U}_s\mathbf{P}_2$ considers perturbation in subspace $\mathbf{U}_s$. Now one needs to determine a set of unknowns matrices $\{P_i\}_{i=1}^4$, $\{Q_i\}_{i=1}^4$ and $\delta\mathbf{\Sigma}_s$ in order to remove all perturbations or refine the series. Let the following assumptions hold according to the SVD of the matrices $\mathbf{X}$ and $\widetilde{\mathbf{X}}$:

$$\mathbf{U}_s^T\mathbf{U}_s = \mathbf{I},\ \mathbf{V}_s^T\mathbf{V}_s = \mathbf{I},\ \mathbf{U}_n^T\mathbf{U}_n = \mathbf{I},\ \mathbf{V}_n^T\mathbf{V}_n = \mathbf{I},\ \mathbf{U}_s^T\mathbf{U}_n = \mathbf{0},\ \mathbf{V}_s^T\mathbf{V}_n = \mathbf{0} \tag{6.2.11}$$

$$\widetilde{\mathbf{U}}_s^T\widetilde{\mathbf{U}}_s = \mathbf{I},\ \widetilde{\mathbf{V}}_s^T\widetilde{\mathbf{V}}_s = \mathbf{I},\ \widetilde{\mathbf{U}}_n^T\widetilde{\mathbf{U}}_n = \mathbf{I},\ \widetilde{\mathbf{V}}_n^T\widetilde{\mathbf{V}}_n = \mathbf{I},\ \widetilde{\mathbf{U}}_s^T\widetilde{\mathbf{U}}_n = \mathbf{0},\ \widetilde{\mathbf{V}}_s^T\widetilde{\mathbf{V}}_n = \mathbf{0} \tag{6.2.12}$$

Let $\mathbf{\Delta}_s = (\mathbf{\Sigma}_s\mathbf{\Sigma}_s^T)^{-1}$ and consider different projections of $\delta\mathbf{X}$ as:

$$\mathbf{E}_{ss} = \mathbf{U}_s^T\delta\mathbf{X}\mathbf{V}_s,\ \ \mathbf{E}_{sn} = \mathbf{U}_s^T\delta\mathbf{X}\mathbf{V}_n,\ \ \mathbf{E}_{ns} = \mathbf{U}_n^T\delta\mathbf{X}\mathbf{V}_s,\ \ \mathbf{E}_{nn} = \mathbf{U}_n^T\delta\mathbf{X}\mathbf{V}_n \tag{6.2.13}$$

Theorems 3

Using the assumptions considered above, the unknowns are, up to second order of $\delta\mathbf{X}$:

$$\mathbf{Q}_1 = \mathbf{\Delta}_s(\mathbf{\Sigma}_s\mathbf{E}_{ss}^T\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{E}_{sn}\mathbf{E}_{nn}^T) + \mathbf{F}_1 \tag{6.2.14}$$

$$\mathbf{Q}_3 = -\mathbf{\Delta}_s\mathbf{E}_{ns}^T\mathbf{E}_{nn} + \mathbf{\Sigma}_s^{-1}\mathbf{E}_{ss}\mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn} + \mathbf{F}_2 \tag{6.2.15}$$

where

$$\mathbf{F}_1 = -\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T, \quad \mathbf{F}_2 = -\mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn} \tag{6.2.16}$$

The other unknowns can be found based on the $\mathbf{Q}_1$, $\mathbf{Q}_3$, $\mathbf{F}_1$ and $\mathbf{F}_2$ as follows:

$$\mathbf{P}_1 = -\mathbf{Q}_1^T, \quad \mathbf{P}_2 = -\tfrac{1}{2}\mathbf{F}_1^T\mathbf{F}_1, \quad \mathbf{P}_3 = -\mathbf{Q}_3^T, \quad \mathbf{P}_4 = -\tfrac{1}{2}\mathbf{F}_2^T\mathbf{F}_2,$$

$$\mathbf{Q}_2 = -\tfrac{1}{2}\mathbf{F}_1^T\mathbf{F}_1, \quad \mathbf{Q}_4 = -\tfrac{1}{2}\mathbf{F}_2^T\mathbf{F}_2,$$

$$\delta\mathbf{\Sigma}_s = \mathbf{E}_{ss} - \mathbf{E}_{sn}\mathbf{F}_2^T - \tfrac{1}{2}\mathbf{\Sigma}_s\mathbf{F}_2\mathbf{F}_2^T + \tfrac{1}{2}\mathbf{F}_1\mathbf{F}_1^T\mathbf{\Sigma}_s \tag{6.2.17}$$

Proof

To proof this, we mainly follow the same procedure considered in [?] with some modifications which is useful for our case. Let us now consider the projections of $\widetilde{\mathbf{X}}$ onto different perturbed subspaces using the assumptions stated in (6.2.12):

$$\tilde{\mathbf{X}}^T\tilde{\mathbf{U}}_s = \tilde{\mathbf{V}}_s\tilde{\mathbf{\Sigma}}_s,\ \tilde{\mathbf{X}}^T\tilde{\mathbf{U}}_n = \tilde{\mathbf{V}}_n\tilde{\mathbf{\Sigma}}_n,\ \tilde{\mathbf{X}}\tilde{\mathbf{V}}_s = \tilde{\mathbf{U}}_s\tilde{\mathbf{\Sigma}}_s,\ \tilde{\mathbf{X}}\tilde{\mathbf{V}}_n = \tilde{\mathbf{U}}_n\tilde{\mathbf{\Sigma}}_n \tag{6.2.18}$$

and in a similar from the projections of $\mathbf{X}$ onto different perturbed subspaces is:

$$\mathbf{X}^T\mathbf{U}_s = \mathbf{V}_s\mathbf{\Sigma}_s,\ \mathbf{X}^T\mathbf{U}_n = \mathbf{V}_n\mathbf{\Sigma}_n,\ \mathbf{X}\mathbf{V}_s = \mathbf{U}_s\mathbf{\Sigma}_s,\ \mathbf{X}\mathbf{V}_n = \mathbf{U}_n\mathbf{\Sigma}_n \tag{6.2.19}$$

Now let us now consider $\tilde{\mathbf{X}}^T\tilde{\mathbf{U}}_s = \tilde{\mathbf{V}}_s\tilde{\mathbf{\Sigma}}_s$.

$$\tilde{\mathbf{X}}^T\tilde{\mathbf{U}}_s = \tilde{\mathbf{V}}_s\tilde{\mathbf{\Sigma}}_s$$

$$(\mathbf{X} + \delta\mathbf{X})^T(\mathbf{U}_s + \delta\mathbf{U}_s) = (\mathbf{V}_s + \delta\mathbf{V}_s)(\mathbf{\Sigma}_s + \delta\mathbf{\Sigma}_s)$$

$$(\mathbf{X} + \delta\mathbf{X})^T(\mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1 + \mathbf{U}_s\mathbf{P}_2) = (\mathbf{V}_s + \mathbf{V}_n\mathbf{P}_3 + \mathbf{V}_s\mathbf{P}_4)(\mathbf{\Sigma}_s + \delta\mathbf{\Sigma}_s) \tag{6.2.20}$$

Equation (6.2.20), using equations (6.2.6) and (6.2.19), is simplified to:

$$\delta\mathbf{X}^T\mathbf{U}_s + \delta\mathbf{X}^T\mathbf{U}_n\mathbf{P}_1 + \delta\mathbf{X}^T\mathbf{U}_s\mathbf{P}_2 + \mathbf{V}_s\mathbf{\Sigma}_s\mathbf{P}_2$$

$$= \mathbf{V}_s\delta\mathbf{\Sigma}_s^T + \mathbf{V}_n\mathbf{P}_3\mathbf{\Sigma}_s^T + \mathbf{V}_n\mathbf{P}_3\delta\mathbf{\Sigma}_s^T + \mathbf{V}_s\mathbf{P}_4\mathbf{\Sigma}_s^T + \mathbf{V}_s\mathbf{P}_4\delta\mathbf{\Sigma}_s^T \tag{6.2.21}$$

Let us now premultiply both sides of (6.2.21) by $\mathbf{V}_n^T$ and $\mathbf{V}_s^T$, respectively, and use the assumption stated in (6.2.11), then we find two new equations which can be useful to find unknowns. Note that these

equation are obtained using $\widetilde{\mathbf{X}}^T\widetilde{\mathbf{U}}_s = \widetilde{\mathbf{V}}_s\widetilde{\mathbf{\Sigma}}_s$:

$$\mathbf{E}_{ss}^T + \mathbf{E}_{ns}^T\mathbf{P}_1 + \mathbf{\Sigma}_s^T\mathbf{P}_2 + \mathbf{E}_{ss}^T\mathbf{P}_2 = \delta\mathbf{\Sigma}_s^T + \mathbf{P}_4\mathbf{\Sigma}_s^T + \mathbf{P}_4\delta\mathbf{\Sigma}_s^T \quad (6.2.22)$$

$$\mathbf{E}_{sn}^T + \mathbf{E}_{nn}^T\mathbf{P}_2 + \mathbf{E}_{sn}^T\mathbf{P}_2 = \mathbf{P}_3\mathbf{\Sigma}_s^T + \mathbf{P}_3\delta\mathbf{\Sigma}_s^T \quad (6.2.23)$$

Similar to those obtained in (6.2.22) and (6.2.23), the following equations can be obtained using other equalities in (6.2.18):

$$\mathbf{E}_{nn}^T + \mathbf{E}_{sn}^T\mathbf{Q}_1 + \mathbf{E}_{nn}^T\mathbf{Q}_2 = \mathbf{0} \quad (6.2.24)$$

$$\mathbf{E}_{ns}^T + \mathbf{E}_{ss}^T\mathbf{Q}_1 + \mathbf{\Sigma}_s^T\mathbf{Q}_1 + \mathbf{E}_{ns}^T\mathbf{Q}_2 = \mathbf{0} \quad (6.2.25)$$

$$\mathbf{E}_{ss} + \mathbf{E}_{sn}\mathbf{P}_3 + \mathbf{\Sigma}_s\mathbf{P}_4 + \mathbf{E}_{ss}\mathbf{P}_4 = \delta\mathbf{\Sigma}_s + \mathbf{P}_2\mathbf{\Sigma}_s + \mathbf{P}_2\delta\mathbf{\Sigma}_s \quad (6.2.26)$$

$$\mathbf{E}_{ns} + \mathbf{E}_{nn}\mathbf{P}_3 + \mathbf{E}_{ns}\mathbf{P}_4 = \mathbf{P}_1\mathbf{\Sigma}_s + \mathbf{P}_1\delta\mathbf{\Sigma}_s \quad (6.2.27)$$

$$\mathbf{E}_{nn} + \mathbf{E}_{ns}\mathbf{Q}_3 + \mathbf{E}_{nn}\mathbf{Q}_4 = \mathbf{0} \quad (6.2.28)$$

$$\mathbf{E}_{sn} + \mathbf{E}_{ss}\mathbf{Q}_3 + \mathbf{\Sigma}_s\mathbf{Q}_3 + \mathbf{E}_{sn}\mathbf{Q}_4 = \mathbf{0} \quad (6.2.29)$$

The unknowns can be obtained using the above equations. It should be noted $\mathbf{Q}_2$ is a Hermitian matrix; $\mathbf{Q}_2 = \mathbf{Q}_2^T$. After some simplifica-

tions the following equality holds between $\mathbf{Q}_1$ and $\mathbf{Q}_2$:

$$\mathbf{Q}_2 \approx \frac{1}{2}\mathbf{Q}_1^T\mathbf{Q}_1 \tag{6.2.30}$$

In addition to the above equality, the following equalities hold:

$$\mathbf{Q}_4 \approx -\frac{1}{2}\mathbf{Q}_3^T\mathbf{Q}_3, \;\; \mathbf{P}_2 \approx -\frac{1}{2}\mathbf{P}_1^T\mathbf{P}_1, \;\; \mathbf{P}_4 \approx -\frac{1}{2}\mathbf{P}_3^T\mathbf{P}_3 \tag{6.2.31}$$

Let us first show that $\mathbf{P}_1 = -\mathbf{Q}_1^T$. To proof this, we need the following Lemma.

Lemma

Let $\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ with SVD's of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ be given in (6.2.6) and (6.2.7), respectively. Assume that $||\delta\mathbf{X}||_2$ is less than the smallest nonzero singular value of $\mathbf{X}$. Let the r dimensional subspace spanned by the columns of $\widetilde{\mathbf{U}}_s$, the perturbed signal subspace, be defined by $\widetilde{\mathbf{S}}_s = span(\widetilde{\mathbf{U}}_s)$ and the $K - r$ dimensional subspace spanned by the columns of $\widetilde{\mathbf{U}}_n$, the perturbed orthogonal subspace, be defined by $\widetilde{\mathbf{S}}_n = span(\widetilde{\mathbf{U}}_n)$. Then, Then $\widetilde{\mathbf{S}}_n$ is spanned by the columns of $\mathbf{U}_n + \mathbf{U}_s\mathbf{Q}_1$ and $\widetilde{\mathbf{S}}_s$ is spanned by the columns of $\mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1$ where $\mathbf{Q}_1$ and $\mathbf{Q}_1$ are matrices whose norms are of the order of $\delta\mathbf{X}$ [?]. The lemma above gives bases for the perturbed signal and orthogonal subspaces. For the orthogonal subspace we have:

$$(\mathbf{U}_n^T + \mathbf{Q}_1^T\mathbf{U}_s^T)(\mathbf{U}_n + \mathbf{Q}_1\mathbf{U}_s) = \mathbf{I} + \mathbf{Q}_1^T\mathbf{Q}_1 \tag{6.2.32}$$

The above equation shows how the basis for the perturbed orthogonal subspace can be normalized. Therefore, an orthonormal basis for the

perturbed orthogonal subspace is given by

$$(\mathbf{U}_n + \mathbf{U}_s\mathbf{Q}_1)(\mathbf{I} + \mathbf{Q}_1^T\mathbf{Q}_1)^{-\frac{1}{2}} \tag{6.2.33}$$

A similar equation holds for the perturbed signal subspace. An orthonormal basis for the perturbed signal subspace is given by

$$(\mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1)(\mathbf{I} + \mathbf{P}_1^T\mathbf{P}_1)^{-\frac{1}{2}} \tag{6.2.34}$$

We know that the perturbed signal and orthogonal subspaces are orthogonal to each other. Thus the unnormalized basis vectors given in the Lemma are orthogonal. That is,

$$(\mathbf{U}_n^T + \mathbf{Q}_1^T\mathbf{U}_s^T)(\mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1) = \mathbf{0} \tag{6.2.35}$$

$$\Rightarrow \mathbf{P}_1 + \mathbf{Q}_1^T = \mathbf{0}, \quad \Rightarrow \mathbf{P}_1 = -\mathbf{Q}_1^T \tag{6.2.36}$$

Therefore, we only need to obtain $\mathbf{Q}_1$ as others can be obtained based on $\mathbf{Q}_1$. Let us now consider $\mathbf{Q}_1$.

The following equality is obtained between $\mathbf{Q}_1$ and $\mathbf{Q}_3$ using (6.2.22)–(6.2.29).

$$\mathbf{E}_{sn} + \mathbf{E}_{ss}\mathbf{Q}_3 + \mathbf{\Sigma}_s\mathbf{Q}_3 = \mathbf{Q}_1\mathbf{E}_{nn} \tag{6.2.37}$$

The above equation can be written as the following form:

$$\mathbf{Q}_3 = \mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn} - \mathbf{\Sigma}_s^{-1}\mathbf{E}_{ss}\mathbf{Q}_3 + \mathbf{\Sigma}_s^{-1}\mathbf{Q}_1\mathbf{E}_{nn} \tag{6.2.38}$$

Now we need to express $\mathbf{Q}_3$ by $\mathbf{Q}_1$. Substituting on the right-hand side

of (6.2.38) and neglecting higher order terms, (6.2.38) is simplified to

$$\mathbf{Q}_3 = \mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn} + \mathbf{\Sigma}_s^{-1}\mathbf{E}_{ss}\mathbf{\Sigma}_s^{-1}\mathbf{E}_{ss} + \mathbf{\Sigma}_s^{-1}\mathbf{Q}_1\mathbf{E}_{nn} \tag{6.2.39}$$

In a similar way, the following equality is obtained between $\mathbf{Q}_1$ and $\mathbf{Q}_3$ using (6.2.22)–(6.2.29):

$$\mathbf{E}_{ns}^T + \mathbf{E}_{ss}^T\mathbf{Q}_1 + \mathbf{\Sigma}_s^T\mathbf{Q}_1 = \mathbf{Q}_3\mathbf{E}_{nn}^T \tag{6.2.40}$$

Substituting $\mathbf{Q}_3$ in (6.2.39) into (6.2.42) and discarding higher order terms, we obtain an equation for $\mathbf{Q}_1$ as follows

$$\mathbf{\Sigma}_s^T\mathbf{Q}_1 = -\mathbf{E}_{ns}^T - \mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn}\mathbf{E}_{nn}^T - \mathbf{E}_{ss}^T\mathbf{Q}_1 \tag{6.2.41}$$

The above equation shows that is not easy to obtain a close form for $\mathbf{Q}_1$ in the current matrix-form equation. However, we can use the recursive technique. Note that we are only interested in the expression of up to the second order of $\delta\mathbf{X}$. Multiplying both sides of (6.2.42) by $\mathbf{\Sigma}_s$, and introducing new definition $\mathbf{\Delta}_s = (\mathbf{\Sigma}_s\mathbf{\Sigma}_s^T)^{-1}$, (6.2.42) becomes

$$\mathbf{Q}_1 \approx -\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{\Delta}_s\mathbf{E}_{sn}\mathbf{E}_{nn}^T - \mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ss}^T\mathbf{Q}_1 \tag{6.2.42}$$

Now, we use recursive method and keeping terms only up to the second-order perturbations, we then use the following matrix form to obtain $\mathbf{Q}_1$,

$$\mathbf{Q}_1 \approx -\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{\Delta}_s\mathbf{E}_{sn}\mathbf{E}_{nn}^T + \mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ss}^T\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T \tag{6.2.43}$$

Now, rearranging all terms in (6.2.43) and new definition $\mathbf{F}_1 = -\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T$, (6.2.43) becomes (6.2.14),

$$\mathbf{Q}_1 = \mathbf{\Delta}_s(\mathbf{\Sigma}_s\mathbf{E}_{ss}^T\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{E}_{sn}\mathbf{E}_{nn}^T) + \mathbf{F}_1 \qquad (6.2.44)$$

6.3 SSA based on the Perturbation Theory

In order to apply the perturbation theory in the SSA technique we need to have a priori information about noise component δY_T or $\delta\mathbf{X}$ (which is the trajectory matrix of the series δY_T). However, the noise series δY_T is unknown in practice and usually there is no a priori information. On way to overcome this problem is to have an estimate of $\delta\mathbf{X}$. Here we use $\widetilde{\mathbf{X}} - \hat{\mathbf{X}}$ as an estimate of $\delta\mathbf{X}$, where $\hat{\mathbf{X}}$ obtained using basic SSA. That is, we first apply the basic SSA technique to the noisy time series to find an initial estimate of $\delta\mathbf{X}$ and then we estimate $\mathbf{X}$ using the perturbation theory approach. Let us now formally describe this algorithm.

Formal description of the proposed technique

Let us have a noise time series $\tilde{Y}_T = (\tilde{y}_1, \ldots, \tilde{y}_T)$. Fix L ($L \leq T/2$), the window length, and let $K = T - L + 1$.

1. (*Computing the trajectory matrix*): transfers a one-dimensional time series $\tilde{Y}_T = (\tilde{y}_1, \ldots, \tilde{y}_T)$ into the multi-dimensional series $\tilde{X}_1, \ldots, \tilde{X}_K$ with vectors $\tilde{X}_i = (\tilde{y}_i, \ldots, \tilde{y}_{i+L-1})^T \in \mathbf{R}^L$, where $K = T - L + 1$. The result of this step is the trajectory matrix $\tilde{\mathbf{X}} = \left[\tilde{X}_1, \ldots, \tilde{X}_K\right]$.

2. (*Constructing a matrix for applying SVD*): compute the matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$.

3. (*SVD of the matrix* $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$): compute the eigenvalues and eigen–vectors of the matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ and represent it in the form $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T = \tilde{P}\tilde{\Lambda}\tilde{P}^T$. Here $\tilde{\Lambda} = diag(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_L)$ is the diagonal matrix of eigenvalues of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ ordered so that $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_L \geq 0$ and $\tilde{P} = (\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_L)$ is the corresponding orthogonal matrix of eigen–vectors of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$.

4. (*Selection of eigen–vectors*): select a group of r $(1 \leq r \leq L)$ eigen–vectors $\tilde{P}_{i_1}, \tilde{P}_{i_2}, \ldots, \tilde{P}_{i_r}$.

   The grouping step corresponds to splitting the elementary matrices $\tilde{\mathbf{X}}_i$ into several groups and summing the matrices within each group.

5. Compute the matrix $\hat{\mathbf{X}} = ||\hat{x}_{i,j}|| = \sum_{k=1}^{r} \tilde{P}_{i_k}\tilde{P}_{i_k}^T\tilde{\mathbf{X}}$.

6. Estimating noise matrix $\delta\mathbf{X}$. To estimate $\delta\mathbf{X}$, we use the difference between the initial estimate of signal matrix $\mathbf{X}$, $\hat{\mathbf{X}}$, and noisy matrix $\widetilde{\mathbf{X}}$; $\delta\mathbf{X} \approx \widetilde{\mathbf{X}} - \hat{\mathbf{X}}$.

7. Estimating signal matrix $\mathbf{X}$ using perturbation theory. An estimation of $\mathbf{X}$ can be reconstructed by perturbation theory; $\mathbf{X} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^T$, where $\mathbf{U}_s$, $\mathbf{\Sigma}_s$ and $\mathbf{V}_s^T$ are refine version of the noisy matrix $\widetilde{\mathbf{U}}_s$, $\widetilde{\mathbf{\Sigma}}_s$ and $\widetilde{\mathbf{V}}_s^T$, respectively, and can be obtained using (6.2.8) and (6.2.10). Note that performing SVD of the estimated noise matrix $\delta\mathbf{X}$ in step 6, enables us to estimate $\mathbf{U}_n$ and $\mathbf{V}_n$.

8. Transition to the one–dimensional series can now be achieved by averaging over the diagonals of the matrix $\mathbf{X}$. Thus, the results

of this step is an approximation of Y_T.

9. The refine series Y_T can now be used for forecasting.

6.4 Empirical results

6.4.1 Simulated data

We shall consider two types of time series; real and artificially generated time series. The capability of the SSA technique based on the perturbation theory (SSA_{PT}), in reconstructing and forecasting, is initially assessed by applying it to the simple sin series:

$$\begin{aligned} S012 &= \beta_0 + \beta_1 Sin(2t\pi/12) + \beta_2 Sin(2t\pi/7) + \beta_3 Sin(2t\pi/5) + \epsilon_t \\ S01 &= \beta_0 + \beta_1 Sin(2t\pi/12) + \beta_2 Sin(2t\pi/7) + \epsilon_t \\ S1 &= \beta_1 Sin(2t\pi/12) + \beta_2 Sin(2t\pi/7) + \epsilon_t \end{aligned} \tag{6.4.1}$$

where ϵ_t is a white noise series. In total 300 data are generated and we added different normally distributed noise to each point of the original series. The simulation was repeated 1000 times. The first 200 observations was considered as in-sample (reconstruction) and the rest as out-of-sample (forecasting). Note that usually every harmonic component with a different frequency produces two eigentriples with close singular values (except for frequency 0.5 which provides one eigentriples with saw-tooth singular vector). For example, one needs to select the first five eigenvalues for reconstruction of the series $S012$, and the first three for the series $S01$. Note also that we need to consider one eigentriple for the intercept, which is the first one in this particular example. Again, to calculate the precision we use the ratio of RMSE (RRMSE).

The effect of window length

Let us first consider the effect of noise reduction with respect to different window length L which is the single parameter in decomposition stage. Certainly, the choice of parameter L depends on the data we have and the analysis we aim to perform. The improper choice of L would imply a inferior decomposition [8]. It should be noted that variations in L may influence separability feature of the SSA technique; the orthogonality and closeness of the singular values. Here we consider L between 10 and 70 which is approximately $T/3$.

Figure 6.1 shows the RRMSE of reconstructed series for different simulated series. As it appears from this figures, SSA_{PT} has a better performance in reconstruction noisy series, particulary for small window length. The performance of both methods are similar for a large window length.

As the figures show, RRMSE tends to 1 as the window length increases confirming that both methods have similar performance for a large window length. The graphs also show that there is a gradual increase in RRMSE with window length. For example for window length 10, the performance of SSA_{PT} is up to 15% better than SSA_{LS} in reconstruction noisy series $S01$. However, there is not a significant discrepancy between the performance of SSA_{PT} and SSA_{LS} for window length greater than 50.

Note that the minimum value of RMSE for both SSA_{PT} and SSA_{LS} occurs for a large window length. Let us, for example, consider the RMSE of SSA_{PT} and SSA_{LS} in reconstructing $S012$ in more details. Figure 6.2 shows the RMSE of SSA_{PT} and SSA_{LS}. As it can be seen from the figure, there is a gradual decrease in RMSE with window

length. In fact, the maximum accuracy in reconstruction, using both methods, occurs for a large window length. The figure also shows that the RMSE of SSA_{PT} is smaller than those obtained using SSA_{LS}. Moreover, the figure indicates that the discrepancy between SSA_{PT} and SSA_{LS} reduces as window length increases. In the rest of this chapter, we only consider the RRMSE as considering two RMSEs and the RRMSE gives equal information, but the RRMSE is more informative.

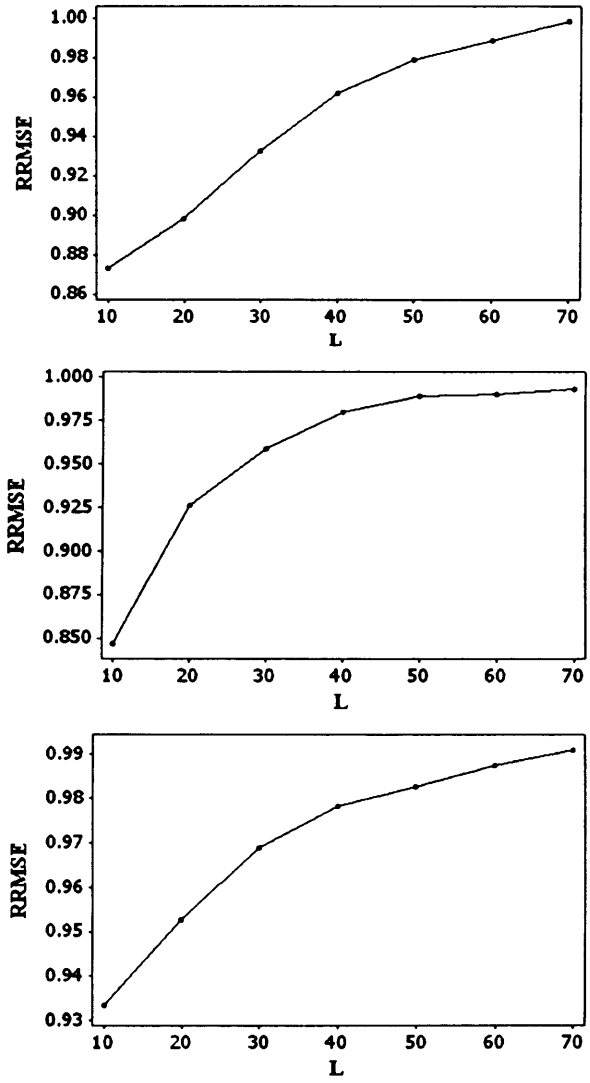


Figure 6.1. The value of RRMSE in reconstructing of noisy series $S012$ (top), $S01$ (middle) and $S1$ (bottom) for different window length.

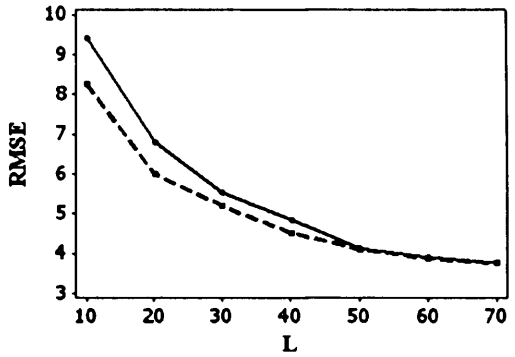


Figure 6.2. The value of RMSE in reconstructing of noisy sin for different window length using SSA_{PT} (dashed line) and SSA_{LS} (thick line).

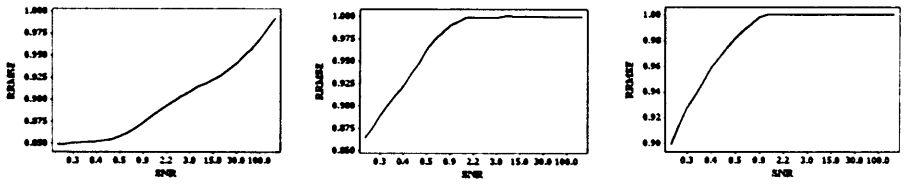
The effect of noise level

To a better understanding the effect of noise reduction with respect to different window length L, we also consider different signal to noise ratio (SNR). Here the SNR is the ratio of standard deviation of the noise free series (signal) to standard deviation of noise. Figures 6.3–6.5 show RRMSE for different values of SNR. For example, Figure 6.3 shows RRMSE for the series $S012$ where we have an intercept and two different harmonic components. As it appears from the figure, there is a gradual increase in RRMSE with SNR. In fact, the minimum RRMSE occurs for a high noise level or lowest SNR. This result confirms that the new SSA algorithm works better for a situation where the series is a mixture of low signal level and high noise level. For example for $L = 10$ and SNR=0.3, the results indicate that the performance of the SSA_{PT} is up to 15% better than the basic SSA_{LS} while this is approximately 4% for SNR=15. However, there is no significant discrepancy between two methods for a series with a high SNR. A similar results can be seen for $L = 40$ and $L = 70$, but the RRMSE tends to 1 faster than for $L = 10$. These results confirm our previous discussion about separability and

window length; lager window length provides better separability.

Let us now consider the problem of separability briefly. For a fix length L, consider a certain SVD of the noisy series $\widetilde{Y}_T$ of length T, and assume that the series $\widetilde{Y}_T$ is a sum of two series Y_T and δY_T; $\widetilde{Y}_T = Y_T + \delta Y_T$. In this case, separability of the series Y_T and δY_T means that we can split the matrix terms of the SVD of the trajectory matrix $\delta\mathbf{X}$ into two different groups, so that the sums of terms within the groups give the trajectory matrices $\mathbf{X}$ and $\delta\mathbf{X}$ of the series Y_T and δY_T, respectively (for more information see [8]).

Figures 6.4 and 6.5 show the results for series $S01$ and $S1$. As the figures show the similar interpretation, as those concluded for series $S012$, can be stated for these series. It should be, however, noted that the RRMSE for more complex series is greater than for a simple series. For example for $L = 10$, the RRMSE is approximately 85% for series $S012$ while this is about 80% and 75% for series $S01$ and $S1$, respectively.

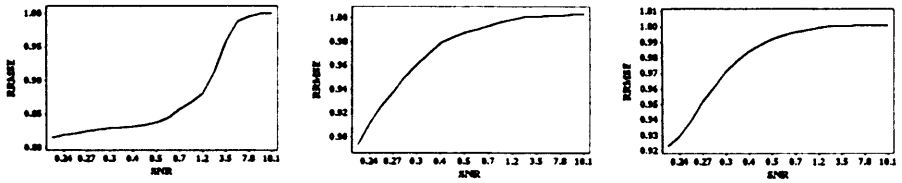


L=10 L=40 L=70

Figure 6.3. The values of RRMES for different noise levels for the series $S012$.

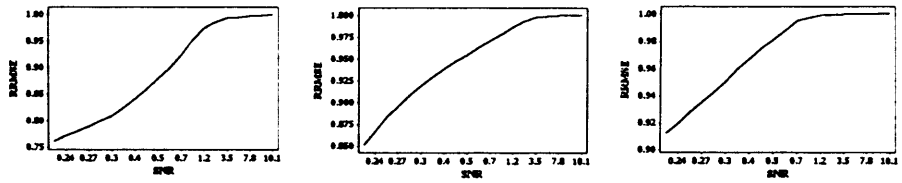
The effect of time-series length

Let us now consider the influence of the time-series length in decomposition and reconstruction of a noisy series. In order to examine this we



L=10 L=40 L=70

Figure 6.4. The values of RRMES for different noise levels for the series $S01$.



L=10 L=40 L=70

Figure 6.5. The values of RRMES for different noise levels for the series $S1$.

used series $S012$, $S01$ and $S1$ with different length N (varies between 100 to 1000). Figure 6.6 shows the value of RRMSE in reconstructing the series $S012$ (thick line), $S01$ (dashed line) and $S1$ (thin line) for different values of N. The results are similar for different values of L and noise levels. As the results show there is no changes in RRMSE as N increases. This is because the series considered here have a structure which can described via a deterministic components. This means the series has a clear structure and this structure is captured well by the SSA. In this context, Hassani et al. [16] showed that in the ideal situation, when we have a series which is a sum of a deterministic component (fully recovered by SSA) and a random noise, the error of the SSA forecast will be exactly the same at all horizons. Here the same results obtained for reconstruction of a series with deterministic com-

ponents. Therefore, we can conclude that for a series which is a sum of a deterministic component and a random noise, the error of the SSA forecast (for h step ahead) and reconstruction (for different series length N) remains stable.

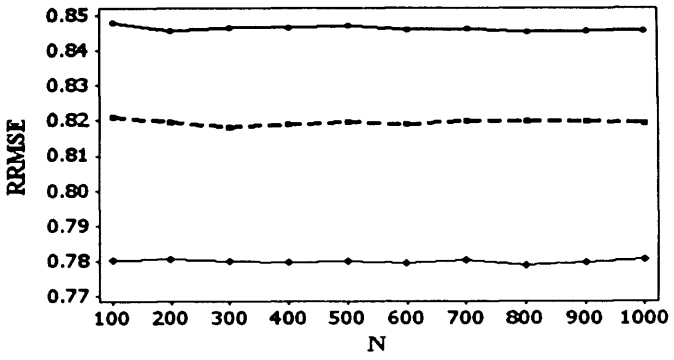


Figure 6.6. The value of RRMSE in reconstructing of noisy series for different N; S012 (thick line), S01 (dashed line) and S1 (thin line).

The effect of Non-stationary noise

So far, we considered the situation where the noise component ϵ_t is stationary. A time series Y_T is called to be stationary if its statistical properties do not depend on time t. Let us now consider the situation where ϵ_t is not stationary. One of the most common instances of non-stationary behaviour is heteroscedasticity, i.e., the variance of noise is proportional to the amplitude of the underlying signal. In the following we examine the capability of SSA_{PT} to detect heteroscedastic noise and reconstructing noise free series. Figure 6.7 (left) shows a realization of the series $S012$ corrupted with a heteroscedasticity noise. Figure 6.7 (right) shows the values of RRMES for different heteroscedasticity noise levels. Here we only represent the results for $L = 10$, but the results are similar for $L = 40$ and $L = 70$ (not shown here). Again, similar to the results obtained for stationary noise, the results indicate that the

performance of SSA_{PT} is much better than those obtained by SSA_{LS}. Therefore, we can conclude that SSA_{PT} works well for detection of a series corrupted with either stationary or non-stationary noise.

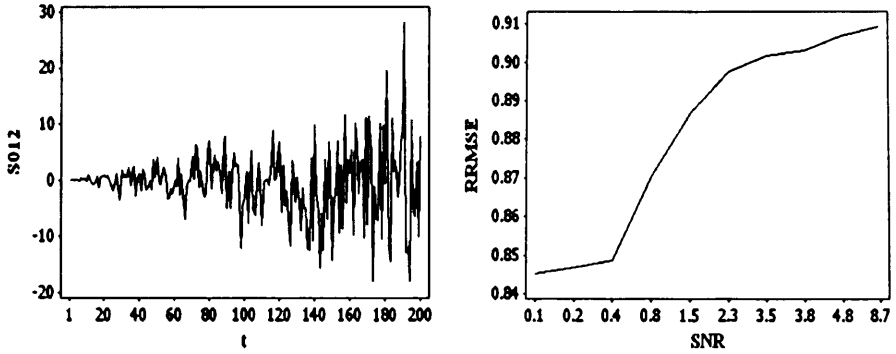


Figure 6.7. Left: A realization of the series $S012$ corrupted with a heteroscedasticity noise. Right: The values of RRMES for different heteroscedasticity noise levels.

6.4.2 Chaotic time series

Next, the capability of the SSA technique as a noise reduction method for chaotic time series was tested by applying the technique to the Hénon map with usual parameter values: $A = 1.4$ and $B = 0.3$ (see Chapter 2). In total 1895 data are generated and we add different normally distributed noise to each point of the original series.

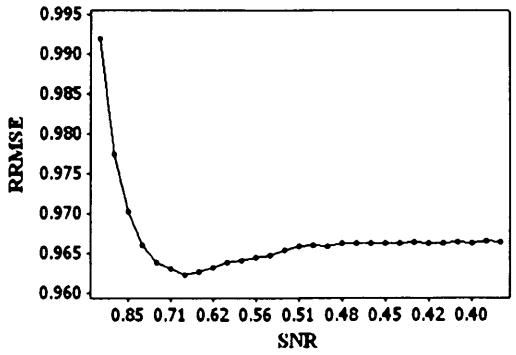


Figure 6.8. The values of RRMES in reconstructing Hénon map.

Figure 6.8 shows the values of RRMES in reconstructing Hénon map for different noise levels. The first two eigenvalues were selected in reconstructing noisy Hénon map. Again, similar to the results obtained for sin series, the results indicate that the performance of SSA_{PT} is slightly better than those obtained by SSA_{LS}. The results indicate that the discrepancy between SSA_{PT} and SSA_{LS} in reconstructing Hénon map is smaller than those obtained for sin series. The performance of SSA_{LS} for filtering of a noisy Hénon map was studied by Hassani et al. [31]. They showed that the SSA_{LS} technique can be used as a powerful noise reduction method for filtering either chaotic series or financial time series. They also showed that the SSA_{LS} performance is much better than considered linear and non-linear models for noisy Hénon map. The new SSA based method represented here can be therefore used as a noise reduction technique for financial time series. It should be noted that if the noise level is higher than the signal level, the SSA_{LS} works better than SSA_{MV} as we remove some parts of the signal component, and consider it as a noise component. In this case, we prefer SSA_{LS}. However, the noise level is usually smaller than the signal level in real case.

6.4.3 Real data

Financial time series.

Hassani et al. [31] considered the daily closing prices of several stock market indices to examine whether noise reduction matters in measuring dependencies of the financial series. Here we also use the same series we used in chapter 3.

Table 6.1 represents a summary of descriptive statistics for the se-

ries before and after filtering. The results in Table 6.1 indicates that the filtered series based on the SSA, for all cases have a smaller standard deviation, S.D, than those values obtained by the GARCH model. Again, as the results shows, the performance of SSA_{PT} is slightly better than SSA_{LS}. The same results can also be seen for the values of the maximum and minimum of the series.

Statistics	Method	DAX 30	CAC 40	FTSE 100	IBEX 35	S&P 500	PSI 20	ASE
	Original	0.24	0.28	0.24	0.36	0.35	0.21	0.58
Mean$\times 10^{-3}$	GARCH	-0.24	-0.21	-0.18	-0.30	-0.17	-0.21	0.84
	SSA_{LS}	0.23	0.28	0.24	0.36	0.35	0.21	0.57
	SSA_{PT}	0.24	0.27	0.25	0.36	0.35	0.21	0.58
	Original	0.11	0.11	0.09	0.11	0.09	0.08	0.16
S.D$\times 10^{-1}$	GARCH	0.11	0.11	0.09	0.11	0.09	0.08	0.16
	SSA_{LS}	0.09	0.09	0.07	0.09	0.09	0.07	0.13
	SSA_{PT}	0.08	0.07	0.06	0.07	0.08	0.07	0.11
	Original	4.57	3.39	3.44	3.83	4.33	8.58	6.94
Kurtosis	GARCH	4.46	3.38	3.44	3.76	4.33	8.44	6.91
	SSA_{LS}	3.81	3.44	3.64	3.77	4.30	6.93	6.10
	SSA_{PT}	3.57	3.31	3.48	3.51	4.12	6.13	5.90

Table 6.1. Descriptive statistics of several stock indices returns series before and after filtering.

	DAX 30	CAC 40	FTSE 100	IBEX 35	S&P 500	PSI 20	ASE
				ACF			
Original	0.0519*	0.0344*	0.0234*	0.0524*	0.0147	0.137*	0.147*
GARCH	0.0001	0.0000	0.0235	-0.0007	0.0145	-0.0001	-0.0001
SSA_{LS}	0.1790*	0.1680*	0.1516*	0.2383*	0.0147	0.4406*	0.4505*
SSA_{PT}	0.1921*	0.1834*	0.1857*	0.2511*	0.0581	0.5437*	0.4728*
				λ			
Original	0.3079*	0.2358*	0.1508*	0.2564*	0.1540*	0.3502*	0.3157*
GARCH	0.2799*	0.1171*	0.1508*	0.5382*	0.1540*	0.7951*	0.2909*
SSA_{LS}	0.2921*	0.2425*	0.2326*	0.2855*	0.1475*	0.4977*	0.5263*
SSA_{PT}	0.3142*	0.2713*	0.2678*	0.2911*	0.1876*	0.5216*	0.5419*

Table 6.2. The values of the ACF at lag-1 and λ of several stock indices returns series before and after filtering.

Table 6.2 shows the values of the ACF at lag-1 and $\lambda = \left(1 - exp[-2I(X,Y)]\right)^{\frac{1}{2}}$ of several stock indices returns series before and after filtering, where $I(X,Y)$ is the mutual information between two series X

and Y. As appears from Table 6.2, the values of the ACF are changed after filtering. In fact, the sign and the direction of ACF, direct and inverse relationship, were changed by filtering. The results indicate that the values of the ACF of the original series and those obtained after filtering by the SSA (except for S&P) are statistically significant. Note that ACF is the cross-correlation of a series with itself. It is the similarity between observations as a function of the time separation between them. The ACF can be regarded as a tool for finding repeating patterns, such as the presence of a harmonic components corrupted with noise term. Note also that ACF at lag-1 is used to detect non-randomness. If random, such autocorrelation should be near zero. For non-random, the autocorrelation will be significantly non-zero.

We also considered the results for for λ. Again, $*$ indicates the results at the 1% level of significance; the values of λ, before and after filtering, are statistically significant.

Monthly accidental deaths in the USA

Below, we examine the performance of the SSA technique based on the perturbation theory by applying it to another real time series, namely, monthly accidental deaths in the USA. The performance of the proposed algorithm were compared with several well-known methods namely, the traditional Box-Jenkins SARIMA models, the ARAR Algorithm and the Seasonal Holt-Winters Algorithm [14].

The results are presented in Table 6.3. The values of RMSE show performance of forecasting. The results confirm that the SSA_{PT} forecasting performance is much better than other forecasting methods. For example for $r = 12$, the value of RMSE for the SSA_{PT} is 40% less

than the first one (model I) and almost 10% less than the ARAR algorithm. Moreover, the results indicate that the SSA_{PT} performance is better than the SSA_{LS}. We also considered the performance of the SSA forecasting results with respect to different values of r. We choose the same window length L but different eigenvalues r (for an explanation of how to choose r and L for this series see chapter 4). The results are presented in Table 6.3, for the first 13, 14 and 15 eigenvalues. As the table shows, again, the SSA_{PT} technique outperforms the other classical methods and also SSA_{PT} is less sensitive than SSA_{LS} for this particular example.

	RRMSE(SSA_{PT}/Other methods)			
Method	$r = 12$	$r = 13$	$r = 14$	$r = 15$
Model I	0.60	0.56	0.59	0.60
Model II	0.62	0.63	0.63	0.65
H-W	0.72	0.68	0.71	0.73
ARAR	0.91	0.86	0.89	0.91
SSA_{LS}	0.98	0.91	0.86	0.74

Table 6.3. The value RRMSE of the post-sample forecasts.

6.5 Conclusion

In this chapter we introduced the SSA technique based on the perturbation theory (SSA_{PT}). The results has illustrated that SSA_{PT} performs well in reconstructing perturbed simulated series. The performance of the proposed algorithm was assessed with respect to different window length L, signal to noise ratio and type of series (stationary and non-stationary). The comparison of the forecasting results showed that SSA_{PT} is much more accurate than several well-known classical methods, in forecasting of a well know time series. We also found that the

SSA_{PT} forecasting results are better than SSA_{LS} for noise reduction of financial time series and chaotic series. In conclusion, the results confirm that both SSA_{PT} and SSA_{LS} give much more accurate results than the classical methods of time series analysis considered here.

Chapter 7

A COMPREHENSIVE CAUSALITY TEST BASED ON THE SINGULAR SPECTRUM ANALYSIS

In this chapter, we consider the concept of causal relationship between two time series based on the singular spectrum analysis. We introduce several criteria which characterize this causality. The criteria are based on the forecasting accuracy and predictability of the direction of change. The performance of the proposed test is examined using different real time series.

7.1 Introduction

A question that frequently arises in time series analysis is whether one economic variable can help in predicting another economic variable. One way to address this question was proposed in [81]. Granger [81] formalized a causality concept as follows: process X does not cause process Y if (and only if) the capability to predict Y series based on the

histories of all observables is unaffected by the omission of X's history (see also [114]). Testing causality, in the Granger sense, involves using F-tests to test whether lagged information on one variable, say X, provides any statistically significant information about another variable, say Y, in the presence of lagged Y. If not, then "Y does not Granger-cause X."

Criteria for Granger causality typically have been realized in the framework of multivariate Gaussian statistics via vector autoregressive (VAR) models. It is worth mentioning that the linear Granger causality is not causality in a broader sense of the word. It just considers linear prediction and time-lagged dependence between two time series. The definition of Granger causality does not mention anything about possible instantaneous correlation between two series X_T and Y_T. (If the innovation to X_T and the innovation to Y_T are correlated then it is sometimes called instantaneous causality.) It is not rare when instantaneous correlation between two time series can be easily revealed, but since the causality can go either way, one usually does not test for instantaneous correlation. In this chapter, several of our causality tests incorporate testing for the instantaneous causality. One more drawback of the Granger causality test is the dependence on the right choice of the conditioning set. In reality one can never be sure that the conditioning set selected is large enough (in short macro-economic series one is forced to choose a low dimension for the VAR model). Moreover, there are special problems with testing for Granger causality in co-integrated relations [127].

The original notion of Granger causality was formulated in terms of linear regression, but there are some nonlinear extensions in the liter-

ature (see, for example, [108]). Hiemstra and Jones [115] also propose a nonparametric test which seems to be most used test in testing nonlinear causality. However, this method also has several drawbacks: *i*) the test is not consistent against a specific class of alternatives [111], *ii*) there are restrictive assumptions in this approach [107] and *iii*) the test can severely over-reject the null hypothesis of non-causality [112].

It is also important to note that Granger causality attempts to capture an interesting aspect of causality, but certainly is not meant to capture all. A method based on the information theory have realized a more general Granger causality measure that accommodates in principle arbitrary statistical processes [110]. Su and White [126] propose a nonparametric test of conditional independence based on the weighted Hellinger distance between the two conditional densities. There are also a number of alternative methods, but they are rarely used.

We overcome all these difficulties by implementing a different technique for capturing the causality; this technique uses the singular spectrum analysis (SSA) technique; a nonparametric technique that works with arbitrary statistical processes, whether linear or nonlinear, stationary or non-stationary, Gaussian or non-Gaussian.

The general aim of this chapter is assessing the degree of association between two arbitrary time series (these associations are often called causal relationships as they might be caused by the genuine causality) based on the observations of these time series. We develop new tests and criteria which is based on the forecasting accuracy and predictability of the direction of change of the SSA algorithms.

7.2 Causality Criteria

7.2.1 Forecasting accuracy based criterion

The first criterion we use here is based on the out-of-sample forecast. The out-of-sample forecast testing is very common in the framework of Granger causality. The question behind Granger causality is whether forecasts of one variable can be improved using the history of another variable. Here, we compare the forecasted value obtained using the univariate procedure, SSA, and also the multivariate one, MSSA. We then compare the predicted values with the actual values to evaluate the forecasting error. If the forecasting error using MSSA is significantly smaller than the forecasting error of the univariate SSA, we then conclude that there is a casual relationship between these series.

Let us consider in more detail the procedure of constructing a vector of forecasting error for an out-of-sample test. In the first step we divide the series $X_T = (x_1, \ldots, x_T)$ into two separate subseries X_R and X_F: $X_T = (X_R, X_F)$ where $X_R = (x_1, \ldots, x_R)$, and $X_F = (x_{R+1}, \ldots, x_T)$. The subseries X_R is used in reconstruction step to provide the noise free series $\tilde{X}_R$. The noise free series $\tilde{X}_R$ is then used for forecasting the subseries X_F using either the recurrent or vector forecasting algorithm formulated above. The subseries X_F will be forecasted using the recursive h-step ahead forecast with SSA and MSSA. The forecasted points $\hat{X}_F = (\hat{x}_{R+1}, \ldots, \hat{x}_T)$ are then used for computing forecasting error. Then the vector $(x_{R+2}, \ldots, x_T)$ is forecasted using the new subseries $(x_1, \ldots, x_{R+1})$ and this procedure is continued recursively up to the end of series, yielding the series of h-step-ahead forecasts for univariate and multivariate algorithms. Therefore, the vector of h-step-ahead

forecast obtained can be used in examining the association (or order h) between the two series. Let us now consider a formal procedure of constructing a criterion of SSA causality of order h between two arbitrary time series.

Criterion

Let $X_T = (x_1, \ldots, x_T)$ and $Y_T = (y_1, \ldots, y_T)$ denote two different time series of length T. Set window length L_x and L_y for the series X_T and Y_T, respectively. Here we assume $L_x = L_y = L$. Using the embedding terminology, we construct trajectory matrices $\mathbf{X} = [X_1, \ldots, X_K]$ and $\mathbf{Y} = [Y_1, \ldots, Y_K]$ for the series X_T and Y_T.

Consider an arbitrary loss function $\mathcal{L}$. In econometrics, the loss function $\mathcal{L}$ is usually selected so that it minimizes the mean square error of the forecast. Let us first assume that the aim is to forecast the series X_T. Thus, the aim is to minimize $\mathcal{L}(X_{K+H_x} - \hat{X}_{K+H_x})$, where vector $\hat{X}_{K+H_x}$ is an estimate, obtained using forecasting algorithm, of the vector X_{K+H} of the trajectory matrices $\mathbf{X}$. Note that, for example, when $H_x = 1$, $\hat{X}_{K+1}$ is an estimate of the vector $X_{K+1} = (x_{T+1}, \ldots, x_{T+h})$ where h varies between 1 and L. In a vector form, this means that an estimate of X_{K+1} can be obtained using the trajectory matrix $\mathbf{X}$ consisting of vectors $[X_1, \ldots, X_K]$. The vector X_{K+H_x} can be forecasted using either univariate SSA or MSSA. Let us first consider the univariate approach. Define

$$\Delta_{X_{K+H_x}} = \mathcal{L}(X_{K+H_x} - \hat{X}_{K+H_x}). \tag{7.2.1}$$

where $\hat{X}_{K+H_x}$ is obtained using univariate SSA; that is, the estimate $\hat{X}_{K+H_x}$ is obtained only from the vectors $[X_1, \ldots, X_K]$.

Let $X_T = (x_1, \ldots, x_T)$ and $Y_{T+d} = (y_1, \ldots, y_{T+d})$ denote two different simultaneous time series and consider the same window length L for both series (here d is the lagged difference between two series). Now, we forecast $x_{T+1}, \ldots, x_{T+h}$ using the information provided by the series Y_{T+d} and X_T. Next, compute the following statistics:

$$\Delta_{X_{K+H_x}|Y_{K+H_y}} = \mathcal{L}(X_{K+H_x} - \tilde{X}_{K+H_x}). \tag{7.2.2}$$

where $\tilde{X}_{K+H_x}$ is an estimate of X_{K+H_x} obtained using multivariate SSA. This means that we simultaneously use vectors $[X_1, \ldots, X_K]$ and $[Y_1, \ldots, Y_{K+H_y}]$ in the forecasting vector X_{K+H_x}. Now, define the following criterion:

$$F_{X|Y}^{(h,d)} = \frac{\Delta_{X_{K+H_x}|Y_{K+H_y}}}{\Delta_{X_{K+H_x}}} \tag{7.2.3}$$

where h indicates h step ahead forecast of the series X_T in presence of the series Y_{T+d} and d shows the lagged difference between series X_T and Y_{T+d}, respectively; here d is any given integer (even negative). For example, $F_{X|Y}^{(h,0)}$ indicates that we use the same series length in forecasting h step ahead series X; we use the series X_T and Y_T simultaneously. $F_{X|Y}^{(h,0)}$ can be considered as a common multivariate forecasting system for the time series with the same series length. The criterion $F_{X|Y}^{(h,0)}$ can then be used in evaluating so-called instantaneous causality. Similarly, $F_{X|Y}^{(h,1)}$ indicates that there is an additional information for series Y and that this information is one step ahead of the information for the series X; we use series X_T and Y_{T+1} simultaneously.

If $F_{X|Y}^{(h,d)}$ is small, then having information of the series Y helps us to have a better forecast of the series X. This means there is a relationship between series X and Y of order h according to this criterion. In

fact, this measure of association shows how much more information about the future values of series X is contained in the bivariate time series (X, Y) than in the series X alone. If $F_{X|Y}^{(h,d)}$ is very small, then the predictions using the multivariate version are much more accurate than the predictions by the univariate SSA. If $F_{X|Y}^{(h,d)} < 1$, then we can conclude that the information provided by the series Y can be regarded as useful or *supportive* for forecasting the series X. Alternatively, if the values of $F_{X|Y}^{(h,d)} \geq 1$, then either there is no detectable association between X and Y or the performance of univariate version is better than multivariate version (this may happen when the series Y has repeated structural breaks which may misdirect the forecasts of X).

To find out which series (X or Y) is more *supportive* in forecasting, we need to consider another criteria. We obtain $F_{Y|X}^{(h,d)}$ in a similar approach. Now, these measures tell us whether using extra information about time series Y_{T+d} (or X_{T+d}) supports in h-step forecasting of X_T (or Y_T). If $F_{Y|X}^{(h,d)} < F_{X|Y}^{(h,d)}$, we then conclude that X is more *supportive* than Y, and if $F_{X|Y}^{(h,d)} < F_{Y|X}^{(h,d)}$, we then conclude that Y is more *supportive* than X.

Let us now consider a definition for a feedback system according to the above criteria. If $F_{Y|X}^{(h,d)} < 1$ and $F_{X|Y}^{(h,d)} < 1$, we then conclude that there is a feedback system between series X and Y. We shall call it F-feedback (forecasting feedback) which means that using multivariate system helps us in forecasting both considered series. We can say that a F-feedback system that X and Y are mutually supportive.

Statistical test

To check if the discrepancy between the two forecasting procedures are statistically significant we may apply the procedure similar to the Diebold and Mariano (1995) test statistic with the corrections suggested by Harvey et al. [69]. The quality of a forecast is to be judged on some specified function $\mathcal{L}$ as a loss function of the forecast error. Then, the null hypothesis of equality of expected forecast performance is $E(D_t) = 0$, where $D_t = (D_{X_{K+H_x}|Y_{K+H_y}} - D_{X_{K+H_x}})$ and $D_{X_{K+H_x}|Y_{K+H_y}}$ and $D_{X_{K+H_x}}$ are the vector of the forecast errors obtained with the univariate and multivariate approaches, respectively. In our case, $\mathcal{L}$ is the quadratic loss function. The modified Diebold and Mariano statistic for h step ahead forecast and the number of n forecasted points is

$$S = \bar{D}\sqrt{\frac{n+1-2h+h(h-1)/n}{n\,\widehat{var}(\bar{D})}}$$

where n is the number of forecasted points, h indicates h step ahead forecast, and $\bar{D}$ is the sample mean of the vector D_t and $\widehat{var}(\bar{D})$ is, asymptotically
$n^{-1}\left(\widehat{\gamma}_0 + 2\sum_{k=1}^{h-1}\widehat{\gamma}_k\right)$, where $\widehat{\gamma}_k$ is the k-th autocovariance of D_t and can be estimated by $n^{-1}\sum_{t=k+1}^{n}(D_t - \bar{D})(D_{t-k} - \bar{D})$. The S statistic follows the asymptotic standard normal distribution under the null hypothesis and its correction for finite sample follows the Student's t distribution with $n-1$ degrees of freedom [69].

7.2.2 Direction of change based criterion

As another measure of forecasting performance, we also compute the percentage of forecasts that correctly predict the direction of change.

Criterion

For the forecasts obtained using only X_T (univariate case), let

$$Z_{X_i} = \begin{cases} 1 & \text{if direction is correct} \\ 0 & \text{Otherwise} \end{cases}$$

for $i = 1, \ldots, n$, where n is the number of forecasted data points. That is, Z_{X_i} takes a value 1 if the forecast series correctly predicts the direction of change and 0 otherwise. $\bar{Z_X} = \sum_{i=1}^{n} Z_{Xi}/n$ shows the proportion of forecasts that correctly predict the direction of the series movement (in forecasting n data points).

For the multivariate case, let $Z_{X|Y}$ takes a value 1 if the forecast series correctly predicts the direction of change of the series X having information about the series Y and 0 otherwise. Then, we define the following criterion:

$$D_{X|Y}^{(h,d)} = \frac{Z_X}{Z_{X|Y}} \tag{7.2.4}$$

where h (h step ahead forecast) and d (lagged difference) have the same interpretation as stated previously for $F_{X|Y}^{(h,d)}$. Therefore, we can obtain $D_{X|Y}^{(h,d)}$ and similarly $D_{Y|X}^{(h,d)}$. The criterion $D_{X|Y}^{(h,d)}$ characterizes the amount of improvement we are getting from the information contained in Y_{T+h} (or X_{T+h}) for forecasting the direction of change in the h step ahead forecast.

If $D_{X|Y}^{(h,d)} < 1$, then having information about the series Y helps us to have a better prediction of the direction of change for the series X. This means that there is an association between the series X and Y with respect to this criterion. In fact this criterion informs us that how much more information we have in the bivariate time series relative

to the information contained in the univariate time series alone with respect to the prediction of the direction of change. Alternatively, if $D_{X|Y}^{(h,d)} > 1$, then the univariate SSA is better than the multivariate version.

To find out which series is more supportive in predicting the direction of change, we consider the following criterion. We obtain $D_{Y|X}^{(h,d)}$ in a similar approach. Now, if $D_{Y|X}^{(h,d)} < D_{X|Y}^{(h,d)}$, then we conclude that that X is more supportive (with respect to predicting the direction) to Y than Y to X.

Similar to the consideration of the forecasting accuracy criteria, we can define a feedback system based on the criteria characterizing the predictability of the direction of change. Let us introduce a definition for a feedback system according to $D_{X|Y}^{(h,d)}$ and $D_{Y|X}^{(h,d)}$. If $D_{Y|X}^{(h,d)} < 1$ and $D_{X|Y}^{(h,d)} < 1$, we conclude that there is a feedback system between the series X and Y for prediction of the direction of change. We shall call this type of feedback D-feedback. Existence of a D-feedback in a system yields that the series in the system help each other to capture the direction of the series movement with higher accuracy.

Statistical test

Let us describe a statistical test for the criterion $D_{X|Y}^{(h,d)}$. As in the comparison of two proportions, when we test the hypothesis about the difference between two proportions, we need first to know whether the two proportions are dependent. The test is different depending on whether the proportions are independent or dependent. In our case, obviously, Z_X and $Z_{X|Y}$ are dependent. We therefore consider this dependence in the following procedure. Let us consider the test statistics for the

difference between Z_X and $Z_{X|Y}$. Assume Z_X and $Z_{X|Y}$, in forecasting n future points of the series X, are arranged as Table 7.1.

$Z_{X\|Y}$	Z_X	number
1	1	a
1	0	b
0	1	c
0	0	d
Total		$n = a + b + c + d$

Table 7.1. An arrangement of Z_X and $Z_{X|Y}$ in forecasting n future points of the series X.

Then the estimated proportion using the multivariate system is $P_{X|Y} = (a + b)/n$, and the estimated proportion using the univariate version is $P_X = (a + c)/n$. The difference between the two estimated proportions is

$$\pi = P_{X|Y} - P_X = \frac{a+b}{n} - \frac{a+c}{n} = \frac{b-c}{n} \tag{7.2.5}$$

Since the two population probabilities are dependent, we cannot use the same approach for estimating the standard error of the difference that is used for independent case. The formula for the estimated standard error for dependent case was given in [113]:

$$\hat{SE}(\pi) = \frac{1}{n}\sqrt{(b+c) - \frac{(b-c)^2}{n}}. \tag{7.2.6}$$

Let us consider the related test for the difference between two dependent proportions, the null and alternative hypotheses

$$\begin{aligned} H_0 &: \pi_d = \Delta_0 \\ H_a &: \pi_d \neq \Delta_0 \end{aligned} \tag{7.2.7}$$

The test statistics, assuming the sample size large enough for normal

approximation to the binomial to be appropriate, is

$$T_{\pi_d} = \frac{\pi - \Delta_0 - 1/n}{\hat{SE}(\pi)} \tag{7.2.8}$$

where $1/n$ is the continuity correction. In our case $\Delta_0 = 0$ or $\pi_d = 0$. That is, the predictability of the direction of change is equal to 50% which is equal to throw a coin. The test statistics then becomes

$$T_{\pi_d} = \frac{(b-c)/n - 1/n}{1/n\sqrt{(b+c) - (b-c)^2/n}} = \frac{b-c-1}{\sqrt{(b+c) - (b-c)^2/n}} \tag{7.2.9}$$

7.3 Comparison with Granger causality test

7.3.1 Linear Granger causality test

Let X_T and Y_T be two stationary time series. To test for Granger causality we compare a full and a restricted model. The full model is given by

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-L+1} + \psi_1 y_{t-1} + \ldots + \psi_p y_{t-L+1} + \varepsilon_{t_{x|y}} \tag{7.3.1}$$

where $\varepsilon_{t_{x|y}}$ is *iid* sequence with zero mean and variance $\sigma_{x|y}$, ϕ_i and ψ_i are model parameters. The null hypothesis stating that Y_T does not Granger cause X_T is

$$H_0 = \psi_{L+1} = \psi_2 = \ldots = \psi_p = 0 \tag{7.3.2}$$

The alternative is at least one $\psi_i \neq 0$ $(i = 1, \ldots, p)$. If the null hypothesis holds, the full model (7.3.1) is reduced to the restricted model as follows:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-L+1} + \varepsilon_{t_x} \qquad (7.3.3)$$

where ε_{t_x} is *iid* sequence with zero mean and variance σ_x. The forecasting results obtained by the restricted model (7.3.3) are compared to those obtained using the full model (7.3.1) to test for Granger causality. We then apply an F-test (or some other similar test) to obtain a p-value for whether the full model results are better than the restricted model. If the full model provides better forecast, according to the standard loss functions, we then conclude Y_T Granger cause X_T. Thus, Y_T would Granger cause X_T if Y_T occurs before and contains information useful in forecasting X_T that is not found in a group of other appropriate variables. As the formula of Granger causality shows the Granger causality test, in fact, is a mathematical formulation which is based on linear regression modeling of two time series. Therefore, the above formulation of Granger causality can only give information about linear features of the series.

Let us now compare similarity and dissimilarity of the proposed algorithm with Granger causality procedure. As was mentioned in the description of the SSA forecasting algorithm the last component y_L of any vector $X = (x_1, \ldots, x_L)^T \in \mathfrak{L}_r$ is a linear combination of the first $L-1$ components $(x_1, \ldots, x_{L-1})$:

$$x_L = \alpha_1 x_{L-1} + \ldots + \alpha_{L-1} x_1.$$

where vector $A = (\alpha_1, \ldots, \alpha_{L-1})$ can be estimated using eigenvectors of the trajectory matrices $\mathbf{X}$. Thus, the univariate version of SSA is

given by

$$x_t = \alpha_1 x_{t-1} + \ldots + \alpha_{L-1} x_{t-L+1} \tag{7.3.4}$$

As can be seen from (7.3.4), a univariate SSA forecasting formula is similar to the restricted model. However, the procedure of parameter estimation in the SSA technique and Granger model are quite different. But both are linear combinations of previous observations. From this point of view, the univariate SSA technique and Granger causality are similar. The multivariate version of SSA is a multivariate system in which we consider both X_T and Y_T simultaneously to estimate vector A. The multivariate forecasting system can be considered as follows:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \alpha_1 x_{t-1} + \ldots + \alpha_{L-1} x_{t-L+1} \\ \beta_1 x_{t-1} + \ldots + \beta_{L-1} x_{t-L+1} \end{pmatrix} \tag{7.3.5}$$

where vectors $A = (\alpha_1, \ldots, \alpha_{L-1})$ and $B = (\beta_1, \ldots, \beta_{L-1})$ are estimated using the multivariate approach. As equation (7.3.5) shows, the multivariate SSA is not similar to the Granger full model. An obvious discrepancy is that we use the value of the series Y in parameter estimation and also in forecasting series X in Granger based test, while we use the information provided in the subspaces generated by Y in multivariate SSA and not the values of observation. More specifically, Granger causality test uses a linear combination of the values of both series X and Y in the full model, whereas multivariate SSA uses the information provided by X and Y in construction of the subspace and not the observations themselves.

7.3.2 Nonlinear Granger causality test

It is worth mentioning that the simultaneous reconstruction of the trajectory matrices $\mathbf{X}$ and $\mathbf{Y}$ in the MSSA technique is also used in testing for Granger causality between two nonlinear time series. Let us consider the concept of nonlinear Granger causality in more detail. Let $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$ be the joint trajectory matrix. In the joint phase space consider a small neighborhood of any vector. The dynamics of this neighborhood can be described via a linear approximation and a linear autoregressive model can be used to predict the dynamics within the neighborhood. Assume the vector of prediction errors are given by $\mathbf{e}_{X|Y}$ and $\mathbf{e}_{Y|X}$. The reconstruction and the fitting procedure are now employed for the individual time series X_T and Y_T in the same neighborhood and the vector of prediction errors $\mathbf{e}_X$ and $\mathbf{e}_Y$ are then computed. Now, we compute the following criteria

$$\frac{Var(\mathbf{e}_{X|Y})}{Var(\mathbf{e}_X)}, \quad \frac{Var(\mathbf{e}_{Y|X})}{Var(\mathbf{e}_Y)} \tag{7.3.6}$$

The above procedure is repeated for various regions on the attractor, each column of trajectory matrices $\mathbf{X}$ and $\mathbf{Y}$, and the average of the above criteria are used. The above criteria, clearly, can be considered as a function of neighborhood size. If the ratios are smaller than 1, we then conclude that there is a nonlinear Granger causal relation between two series. The similarity of nonlinear Granger causality test with SSA causality test is only in the constructing of trajectory matrices $\mathbf{X}$ and $\mathbf{Y}$ using embedding terminology which is only the first step of SSA. Otherwise, the Granger nonlinear test is totally different from the test considered here. Moreover, the major drawback of the standard

nonlinear analysis is that it requires a long time series, while the SSA technique works very well for short and long time series (for example see [12]).

7.3.3 More about the dissimilarity between Granger causality and the SSA-based techniques

Let us return to the discussion about dissimilarities between the Granger causality and the SSA-based causality. One of the main drawbacks of the Granger causality is we need to assume that the model is fixed (we then just test for a significance of some parameters in the model). However, the model can be (and usually is) wrong. The test statistics used for testing the Granger causality are not comprehensive. In the main case of linear model, testing for Granger causality consists in the repeated use of the standard F-test which is famously sensitive with respect to various deviations from the model. The Granger causality is only associated with lag difference between the two series.

In the approach we develop in this chapter, the model of dependence (or causality) is not fixed a priori; instead, this model is built in the process of analysis. The models we build are non-parametric and are very broad (in particular, causality is not necessarily associated with lag) and flexible.

The tests for Granger causality consider the past information of other series in forecasting the series. For example, as we mentioned in linear Granger causality test, we use the series X up to time t and the series Y up to time $t-d$; the series Y_{T-d} is used in forecasting series X_T. Whereas in the proposed test here, the series Y_{T+d} is employed in forecasting series X_T.

Furthermore, the tests for Granger causality are based on the forecasting accuracy. Here, we have also introduced another criterion for capturing causality which is based on the predictability of the direction of change. As we mentioned above for some purposes, it may be more harmful to make a smaller prediction error yet fail in predicting the direction of change, than to make a larger directionally correct error [106].

Moreover, the definition of Granger causality does not mention anything about possible instantaneous correlation between two series X_T and Y_T. Recall that if the innovation to X_T and the innovation to Y_T are correlated we say there is instantaneous causality. The criteria we introduced here enables us to have an interpretation for instantaneous causality. In fact, the proposed test is not restricted on lagged difference between two series. It works even when there is no lagged difference between series.

Furthermore, real world time series (e.g., financial time series) are typically noisy, non-stationary, and can have small length. It is well known that the existence of a significant noise level reduces the efficiency of the tests (linear and nonlinear) for Granger causality.

There are mainly two different approaches to examine causality between two time series. According to the first one, that is utilized in current methods, the criteria of capturing causality is computed directly from the noisy time series. Therefore, we ignore the existence of the noise in the first approach. This can lead to misleading interpretations of causal effects. According to the second approach, which we are using in the proposed test, we start with filtering the noisy time series in order to reduce the noise level and then calculate the criteria. It

is commonly accepted that the second approach is more effective than the first one if we are dealing with the series with high noise level.

7.4 Index of Industrial Production Series

Let us now consider the index of industrial production (IIP) series. The IIP series is a key indicator of the state of the UK's industrial base and regarded as a leading indicator of the general state of the economy. The IIP series is published on a monthly basis by the Office for National Statistics (ONS). The index is first released as a provisional estimate and then revised each month to incorporate the information that was not available at the time of the preliminary release. A number of studies have been concerned with the size and nature of revisions to important economic time series. Patterson and Heravi [123–125] have extensively analyzed the key national income and expenditure time series. There are many other studies for modelling and forecasting of data revision. For example, Patterson [120, 121] have used state space approach in forecasting the final vintage of the IIP series and real personal disposable income. For more information about the data revision see [119, 122, 125].

The overall data period for the study includes 423 monthly observations for 1972:1 to 2007:3 on 12 vintages of data seasonally adjusted IIP. The first vintage, which is published one month after the latest month of published data, refers to the first publication in the monthly Digest of statistics. The second vintage refers to the next published figure and so on. For this study we take the 12^{th} vintage as the final vintage (m), then having 12 vintages of data on the same variables.

Let y_t^v be the vth vintage ($v = 1, \cdots, m$) of the data on variable

y for the period t, where $v = 1$ indicates the initially published data and $v = m$ the finally published data. (In practice, m may be taken to indicate the conditionally final vintage.) Here $m = 12$. The structure of the data which is published by *Monthly Digest of Statistics* (MDS) is as follow:

$$\begin{pmatrix} y_1^1 & y_1^2 & y_1^3 & \cdots & y_1^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{t-m}^1 & y_{t-m}^2 & y_{t-m}^3 & \cdots & y_{t-m}^m \\ \vdots & \vdots & \vdots & \ddots & \\ y_{t-2}^1 & y_{t-2}^2 & y_{t-2}^3 & & \\ y_{t-1}^1 & y_{t-1}^2 & & & \\ y_t^1 & & & & \end{pmatrix}. \tag{7.4.1}$$

Thus, publication from a particular issue of MDS traces back a diagonal of this data matrix which is a composite of data of different vintages. We expect that there is a SSA causal relationship between preliminary vintage (v^{th} vintage) and final vintage (m^{th} vintage). To answer this, we need to forecast h step ahead ($h = 1, \ldots, 11$) of the final vintage, $v = m$, giving the information at time t. The forecast could be obtained using classical univariate time series methods. However, the forecasts are not optimal since other information (vintages) available at time t are not used. For example, in forecasting y_{t-m+1}^m we also have available information of y_{t-m+1}^v for $v = 1, \cdots, m-1$, each of which could itself be regarded as a forecast of y_{t-m+1}^m. This matter motivates us to use multivariate method for forecasting h step ahead of y_t^m. For example, to obtain the final vintage value at time t, y_t^m, we can use the information for the first vintage data $y_1^1, \ldots, y_t^1$ and the

final vintage data $y_1^m, \ldots, y_{t-m}^m$. If the results of h step ahead forecast MSSA are better than SSA, e.g. $F_{v^m|v^i}^{(h,m-i)} < 1$ and $D_{v^m|v^i}^{(h,m-i)} < 1$, we then conclude that there is a SSA causal relationship of order h between i^{th} vintage and final vintage. To find out this, SSA and MSSA models are estimated using data to the end of 2000 and post-sample forecasts are then computed for 64 observations of 2001:1-2006:3. Thus, we have 64 one step ahead post sample forecast errors, at horizon $h = 1$. The number of forecast errors available decreases as the forecast horizon increases, so that at horizons of $h = 2, 3, \ldots, 12$ the number of forecast errors are $63, 62, \ldots, 52$ respectively. The value of $F_{v^m|v^i}^{(h,m-i)}$ and $D_{v^m|v^i}^{(h,m-i)}$ $(i = 1, \ldots, 11)$ for each vintage and relative to single SSA are given in Table 7.2. The two parameters L (window length) and r (number of eigenvalues) chosen in the decomposition and reconstruction are also presented in the table.

As it appears from Table 7.2, there are gains to using MSSA throughout the revision process, these being between 87% and 67% for vintage up to $v = 5$, reducing to 50% or slightly less for latter vintages (according to the column labeled $F_{v^m|v^i}^{(h,m-i)}$). This is because, as the structure of the data matrix (7.4.1) shows, even one observation is very important in forecasting a new vector of the data matrix (7.4.1). All results are statistically significant at the 1% significant level.

For the direction of change results, for each preliminary vintage v, we compare the true direction of $y_t^m - y_{t+v-12}^m$ with the direction of vintage v estimate $y_t^v - y_{t+v-12}^m$ and the SSA estimate $\hat{y}_t - y_{t+v-12}^m$. Table 7.2 provides the percentage of forecasts that correctly predict the direction of change for each vintage. As the results show the percentage of correct signs produced by MSSA are significantly higher than those

given by SSA, these being between 55% and 45% for vintage up to $v = 5$, reducing to 18% for latter vintages (according to the column labeled $D^{(h,m-i)}_{v^m|v^i}$).

Thus, these results, without exception, confirm that there exist the SSA causal relationship between each vintage and final vintage. In fact the results with strong evidence indicate that the SSA causal between i^{th} vintage and final vintage is of order $m - i$. It should be noted that here i is equal to h step ahead forecast which is the time lag difference between i^{th} vintage and final vintage. Here, as the results show, the SSA causality holds for lower order such as the results we found for exchange rate series. This confirms that SSA causality of order $m - i$ consequences other order of causality. Note that here the problem of interest is one side causality as we only forecast the final vintage.

Note also that, again the results of Granger causality test, shows that there is a Granger causal relationship between these series. This is not surprising as each column of the data matrix is a revised version of the previous column and therefore they are high correlated. Also, it should be noted that the results of VAR model in forecasting these series are worse than the MSSA results. As the aim of this research is not forecasting, we do not provide the forecasting results here.

7.5 Conclusion

In this chapter, we developed a new approach in testing for causality between two arbitrary univariate time series. We introduced a family of causality tests which are based on the singular spectrum analysis (SSA) analysis. The SSA technique accommodates, in principle, arbitrary statistical processes, whether linear, nonlinear, stationary, non-stationary,

i^{th} Vintage	L	r	$F_{v^m\|v^i}^{(h,m-i)}$	$D_{v^m\|v^i}^{(h,m-i)}$
1	13	5	0.22*	0.45*
2	12	5	0.24*	0.47*
3	11	5	0.27*	0.48*
4	10	5	0.31*	0.50*
5	9	5	0.33*	0.55*
6	8	4	0.36*	0.61*
7	7	4	0.39*	0.65*
8	6	3	0.41*	0.70*
9	5	3	0.45*	0.73*
10	4	3	0.49*	0.77*
11	3	2	0.55*	0.82

Table 7.2. The value of $F_{v^m|v^i}^{(h,m-i)}$ and $D_{v^m|v^i}^{(h,m-i)}$ in forecasting of i^{th} vintage of the index of industrial production series.

Gaussian, or non-Gaussian. Accordingly, we believe our approach to be superior to the traditional criteria used in Granger causality tests, criteria that are based on autoregressive moving average (p, d, q) or multivariate vector autoregressive (VAR) representation of the data; the models that impose restrictive assumptions on the time series under investigation.

Several metrics and criteria are introduced in testing for casuality. The criteria are based on the idea of minimizing a loss function, forecasting accuracy and predictability of the direction of change. We use the univariate SSA and multivariate SSA in forecasting the value of the series and also prediction of the direction.

The performance of the proposed test was examined using the index of industrial production (IIP) series for the United Kingdom. Moreover, it has been documented that, without exception, there exists a SSA causal relationship between each vintage and final vintage of the IIP data.

Chapter 8

SUMMARY AND CONCLUSION

Given that the dynamics of the economy of many countries has gone through many policy and structural changes over different periods of time, one needs to make certain that the method of prediction is not sensitive to the dynamical variations.

The SSA method is highly adaptive in determining the principal features of a nonstationary time series process because it uses density functions derived from the singular value decomposition (SVD) singular vectors to generate moments that are associated with the principal features of the nonstationary process.

It should be noted that in the SSA many probabilistic and statistical concepts are employed, however, the technique is non-parametric and does not make any statistical assumptions such as stationarity concerning either signal or noise in the data. One may consider this as one of the advantages of the technique compared to other classical methods which usually rely on some restricted assumptions.

In this research, we have described the methodology of SSA and demonstrated that SSA can be successfully applied to the analysis and forecasting of economic time series. This research has illustrated that

the SSA technique performs well in the simultaneous extraction of harmonics and trend components. The comparison of forecasting results showed that SSA is more accurate than several well-known methods, in the analysis and future prediction of the several economics time series. The series considered in this research are some examples of different seemingly complex series with potential structure which can be easily analysed by SSA and could provide a typical example of a successful application of SSA.

For example, we compared SSA, ARIMA and Holt-Winter methods for forecasting seasonally unadjusted monthly data on industrial production indicators in Germany, France and the UK. The results have demonstrated that SSA is a very powerful tool for analyzing and predicting economic data. SSA outperformed the ARIMA and Holt-Winter methods in predicting the values of the production series according to the RMSE criterion, particularly at long horizons. The SSA technique outperforms the Holt-Winter and ARIMA models at longer horizons and hence can be considered as a reliable method for predicting recessions and expansions.

The results also show that SSA works well for small sample sizes, as for the UK with the sample size of 84 observations. The forecasts obtained by bootstrapping also confirm the findings.

Moreover, to analysis even more short time series, I have used 32 Iranian national account data sets describing the main economic features of the Islamic Republic of Iran. The data are given in a quarterly and yearly format and have different types of non-stationarity. All the data sets are rather short.

The results show that SSA can be successfully used for the anal-

ysis and forecasting of short economic time series with different types of non-stationarity. In particular, many quarterly series have periodic components with non-stationary amplitudes but SSA has been able to extract and forecast these periodic components very accurately. Most of the yearly data have clear structural changes which makes the application of standard methods of analysis almost impossible.

Another finding, which is very important in forecasting economic time series, is that unlike standard methods used for analysis of economics time series, SSA does not require parametric models or transformation of the data into the logarithmic scale. Moreover, our study has shown that in most cases, the transformation of the quarterly series into logarithmic scale has lead to the deterioration of the precision of the forecasts.

The univariate and multivariate SSA was used in prediction of value and direction of changes (series moving up or down) in the daily UK exchange rates. The empirical results and the test statistics show that MSSA have outperformed random walk models for the pound / dollar exchange rate series (similar results were obtained for the euro/dollar series. The results of unit root test indicated that both UK and EU series are non-stationary series. The results of cointgeration also confirmed that the series are cointegrated. The error correction (EC) model for the cointgerated series was used for prediction. The prediction results based on EC model show an inferior performance compared to predictions by a RW as well as SSA and MSSA methods. We perfo rmed Granger causality test and found that there exists a two-way causality between pound/dollar and EU/dollar exchange rates.

Given that the traditional structural econometric models of ex-

change rates have a poor record in prediction of the exchange rates in comparison to random walk models, we believe SSA and MSSA methods are highly promising. As is shown in this thesis, the SSA method, multivariate representation, has decisively outperformed random walk models for exchange rate series. Further methodological development in this field as well as extensive application of these methods in financial and economic data could prove to be indispensable for accurate prediction exercises.

We have also utilized several price indexes including consumer price index with and without highly volatile food and energy items as well as quarterly Chain-weighted GDP and GNP price indexes for forecasting inflation rate and price levels.

The results show that the SSA significantly outperforms all other methods commonly used in inflation forecasting. I believe the superior prediction results are based on the capability of the SSA method to discard the stochastic components of the original series.

The results show that without exception, SSA outperforms both the naive random walk method and more complex econometric models that are used by other researchers in forecasting inflation rate based on the GDP price index. Moreover, we find that MSSA outperforms the random walk predictions in both one and 3-step ahead forecasts as well as all other time periods considered for forecasting inflation rate based on the CPI-all and CPI-core series. We also find that SSA performs very well in predicting the direction of change.

We also compared the MSSA forecasting results with those results obtained by Phillips curve, DFM and AR(p) models. Once again, MSSA outperforms all other models for forecasting inflation rate and

direction of change in the CPI-all and CPI-core.

In light of inadequate performances of the NAIRU Philips curve-based and the time series models, we conclude that using SSA and MSSA is more promising for obtaining accurate forecasting of inflation rate.

Finally, we developed a new approach in testing for causality between two arbitrary univariate time series. We introduced a family of causality tests which are based on the SSA technique.

Several metrics and criteria are introduced in testing for casuality. The criteria are based on the idea of minimizing a loss function, forecasting accuracy and predictability of the direction of change. We use the univariate SSA and multivariate SSA in forecasting the value of the series and also prediction of the direction.

The performance of the proposed test was examined using the euro/dollar and the pound/dollar daily exchange rates as well as the index of industrial production (IIP) series for the United Kingdom. It has been shown here that the euro/dollar rate causes the pound/dollar rate and vice versa. Moreover, it has been documented that, without exception, there exists a SSA causal relationship between each vintage and final vintage of the IIP data.

The SSA technique accommodates, in principle, arbitrary statistical processes, whether linear, nonlinear, stationary, non-stationary, Gaussian, or non-Gaussian. Accordingly, we believe our approach to be superior to the traditional criteria used in Granger causality tests, criteria that are based on autoregressive moving average (p, d, q) or multivariate vector autoregressive (VAR) representation of the data; the models that impose restrictive assumptions on the time series un-

der investigation.

Appendix A

MEASURES OF ACCURACY AND STATISTICAL SIGNIFICANCE OF THE PREDICTIONS

To measure the performance of the methods of prediction time series, the root mean square error (RMSE) and mean relative absolute error (MRAE) is used. The RMSE is the most frequently quoted measure in forecasting literature [128]. To make sure that the SSA results are not chance occurrence, the modified Diebold-Marino test statistics is used. Additionally, the direction of change criterion is employed which shows the proportion of forecasts that correctly predict the direction of the movement of the series.

A.1 Root mean square of errors (RMSE)

As a measure of prediction accuracy, the following ratio of root-mean-square errors (RMSE) is used:

$$\text{RMSE} = \left(\frac{\sum_{i=1}^{n} (y_{T+i} - \widehat{y}_{T+i})^2}{\sum_{i=1}^{n} (y_{T+i} - \widetilde{y}_{T+i})^2} \right)^{1/2}.$$

Here n represents the number of forecasted points, $\widehat{y}_{T+i}$ are the forecasted values of y_{T+i} obtained by SSA and $\widetilde{y}_{T+i}$ is the forecasted values of y_{T+i} obtained by other method. Note that $\widetilde{y}_{T+i}$ for Random walk (RW) model is y_{T+i-h} for any h-step ahead forecasting. If RMSE < 1, then SSA procedure outperforms alternative prediction method. Alternatively, RMSE > 1 would indicate that the performance of the corresponding SSA procedure is worse than the predictions of the competing method.

A.2 Diebold-Marino significance test

As stated above, to check if the differences between the two forecasting procedures are statistically significant we applied the Diebold and Mariano (1995) test statistic with the corrections suggested by Harvey et al. [69]. The quality of a forecast is to be judged on some specified function $g(e)$ as a loss function of the forecast error, e. Then, the null hypothesis of equality of expected forecast performance is $E(d_t) = 0$, where $d_t = [g(e_{SSA}) - g(e_{RW})]$ and e_{SSA} and e_{RW} are the forecast errors obtained with SSA and RW model, or the other methods, respectively. In our case, g is the quadratic loss function. The Diebold and Mariano statistic for h step ahead forecast and the number of n forecasted points

is

$$S = \bar{d}\sqrt{\frac{n+1-2h+h(h-1)/n}{n\,\widehat{var}(\bar{d})}}$$

where $\bar{d}$ is the sample mean of the d_t series and $\widehat{var}(\bar{d})$ is, asymptotically $n^{-1}\left(\widehat{\gamma}_0 + 2\sum_{k=1}^{h-1}\widehat{\gamma}_k\right)$, where $\widehat{\gamma}_k$ is the k-th autocovariance of d_t and can be estimated by $n^{-1}\sum_{t=k+1}^{n}(d_t - \bar{d})(d_{t-k} - \bar{d})$. The S statistic follows the asymptotic standard normal distribution under the null hypothesis and its correction for finite sample follows the Student's t distribution with $n-1$ degrees of freedom.

A.3 Mean Relative Absolute Error (MRAE)

There are a number of proportional measures that can also be used for description of relative error of the series. The mean absolute percentage error measure the relative amount of error or bias in the forecast. The mean absolute relative error is as follows:

$$\text{MRAE} = \frac{\sum_{i=1}^{n}(|y_{T+i} - \widehat{y}_{T+i}|)}{\sum_{i=1}^{n}(|y_{T+i} - \widetilde{y}_{T+i}|)}.$$

where n represents the number of forecasted points, $\widehat{y}_{T+i}$ are the forecasted values of y_{T+i} obtained by SSA and $\widetilde{y}_{T+i}$ is the forecasted values of y_{T+i} obtained by other method.

If MRAE < 1, then SSA procedure outperforms alternative prediction method. Alternatively, MRAE > 1 would indicate that the performance of the corresponding SSA procedure is worse than the predictions of the competing method.

A.4 Direction of change criterion

The third characteristic computed for each method is the direction of change criterion (DC). It shows the proportion of forecasts that correctly predict the direction of the series movement. Let Z_t $(t = T+1, \ldots, T+n)$ takes a value 1 if the forecast series correctly predicts the direction of change and 0 otherwise. The Moivre-Laplace central limit theorem implies that for large samples the test statistic $2(\bar{Z} - 0.5)n^{1/2}$ is approximately distributed as standard normal. When $\bar{Z} = \sum_{t=1}^{n} Z_t/n$ is significantly larger than 0.5, the forecast is said to have the ability to predict the direction of change. Alternatively, if $\bar{Z}$ is significantly smaller than 0.5, the forecast tends to give the wrong direction of change.

Appendix B

FILTERING METHODS

B.0.1 Autoregressive Moving Average: ARMA

For a large class of autocovariance functions $\gamma(h)$ it is possible to find an ARMA process Y_T with autocovarince function $\gamma_y(h)$ such that $\gamma(h)$ is well approximated by $\gamma_y(h)$. In particular, for any positive integer k, there exists an ARMA process Y_T such that $\gamma_y(h) = \gamma(h)$ for $k = 0, \ldots, h$. For this reason the family of ARMA processes plays a key role in the modeling of dependent data.

The stationary time series Y_T is an ARMA(p, q) process if for every t, $\Phi(B)y_t = \Theta(B)z_t$, where z_t is a white noise process with mean zero and variance σ^2, $\Phi(B) = 1 - \phi_1 B_1 - \ldots - \phi_p B_p$, $\Theta(B) = 1 - \theta_1 B_1 - \ldots - \theta_q B_q$ and B is the backward shift operator defined by $B_j(y_t) = y_{t-j}$. Detailed discussions of the method can be found in Brockwell and Davis [42]. The use of the ARMA model as a benchmark model should not imply that we believe the model is an optimal filtering method for financial series. We use this model as a linear and benchmark model.

B.0.2 Generalized Autoregressive Conditional Heteroskedasticity: GARCH

Autoregressive conditional heteroskedasticity models, introduced by Engle [129] and later generalized by Bollerslev [130], are widely used in various financial applications such as risk management, option pricing, foreign exchange, and the term structure of interest rates [131]– [133]. They explicitly parameterize the time-varying volatility in terms of past conditional variances and past squared innovations (prediction errors), while taking into account excess kurtosis (i.e., heavy tail behavior) and volatility clustering, two important characteristics of financial time-series.

Let y_t denote a real-valued discrete-time stochastic process, and ψ_t denote the information set available at time t. Then, the prediction error ε_t at time t regarding to minimum mean-squared error is obtained as $\varepsilon_t = y_t - E(y_t|\psi_{t-1})$. The conditional variance of y_t given the information through time $t-1$ is by definition the conditional expectation of ε_t^2; $\sigma_t^2 = var(y_t|\psi_{t-1})$.

Let z_t be a zero-mean unit-variance white noise process with some specified probability distribution. Then a GARCH model of order (p,q), denoted by $\varepsilon_t \sim GARCH(p,q)$, has the following general form:

$$\varepsilon_t = \sigma_t z_t \tag{B.0.1}$$

$$\sigma_t = \left(f(\sigma_{t-1}^2, \ldots, \sigma_{t-p}^2, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2)\right)^{\frac{1}{2}} \tag{B.0.2}$$

That is, the conditional variance σ_t^2 is determined by the values of p past conditional variances and q past squared innovations, and the predictive error ε_t is generated by scaling a white noise sample with the

conditional standard deviation. The most widely used GARCH model specifies a linear function f in Eq.(B.0.2) as follows:

$$\sigma_t^2 = c + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_j \varepsilon_{t-j}^2 \tag{B.0.3}$$

where, $c > 0, \alpha_i \geq 0, \beta_j \geq 0$, $(i = 1, \ldots, q,\ j = 1, \ldots, p)$ and $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$.

Appendix C

LINEAR AND NONLINEAR MEASURES OF DEPENDENCE

C.1 Linear correlation coefficient and autocorrelation

Linear correlation is generally used to measure the linear association between two variables. The linear correlation coefficient, ρ, between two random variables X and Y is defined as:

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \qquad \text{(C.1.1)}$$

where E is the expected value operator, μ_X, σ_X and μ_Y, σ_Y are expected value and standard deviation of random variables X and Y, respectively. The sample linear correlation coefficient of T observations of random variables X and Y, can be obtained by replacing μ_X and μ_Y with the sample mean $\bar{x}$ and $\bar{y}$ and also σ_X and σ_Y with the sample standard deviations s_x and s_y (as estimators of μ_X, μ_Y and σ_X, σ_Y) in Eq.(C.1.1), respectively.

Pearson's correlation coefficient, ρ, has the advantage of being a

real-number easy to compute and to interpret. However, it suffers from a large number of drawbacks as follows:

i) it only detects linear dependencies in data; non-linear patterns, even simple ones, can not be measured,

ii) it is only defined when the variance is finite,

iii) it is not a distribution-free measure. It describes completely the dependence structure in a normal population. However, it is now well-known and empirically proved that the Gaussian framework does not describe reality, especially due to the presence of heavy-tails in empirical financial distributions,

iv) it is not invariant under non-linear strictly increasing transformations.

As it appears from Eq.(C.1.1), we clearly see that ρ is highly influenced by the variance. Hence, even few extreme observations can imply a high variance in the denominator, and therefore, can bias the correlation coefficient.

The autocorrelation function (ACF) of a stationary time series Y_t at lag h is:

$$\frac{Cov(Y_{t+h}, Y_t)}{\sigma_Y^2} = \frac{\gamma(h)}{\gamma(0)} \tag{C.1.2}$$

In practical problems of course we only have a set of data $Y_T = (y_1, \ldots, y_T)$. Therefore, the sample autocovariance function, $\hat{\gamma}(h)$, is defined as

$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{T-h}(y_{t+h} - \bar{y})(y_t - \bar{y})}{T} \tag{C.1.3}$$

and then the sample ACF at lag h is $\hat{\gamma}(h) \,/\, \hat{\gamma}(0)$.

C.2 Mutual information

The mutual information of two continuous random variables X and Y can be defined as:

$$I(X;Y) = \int_X \int_Y P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right) d_y d_x \qquad \text{(C.2.1)}$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y, respectively. In the discrete case, we replace the integral by a definite double summation. Intuitively, mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces our uncertainty about the other. Mutual information can be expressed as:

$$I(X;Y) = H(X)-H(X|Y) = H(Y)-H(Y|X) = H(X)+H(Y)-H(X,Y) \qquad \text{(C.2.2)}$$

where $H(X)$ and $H(Y)$ are the marginal entropies, $H(X|Y)$ and $H(Y|X)$ are the conditional entropies, and $H(X,Y)$ is the joint entropy of X and Y.

Since $H(X) \geqslant H(X|Y)$, we have $I(X;Y) \geqslant 0$; assuming equality iff X and Y are statistically independent. Therefore, the mutual information between the vectors of random variables X and Y can be considered as a measure of dependence between these variables, or better yet, the statistical correlation of X and Y. The statistics defined in Eq.(C.2.2) satisfies some of the desirable properties of a good measure of dependence [45].

The main difficulty in estimating the mutual information from the

empirical data lies in the fact that the relevant probability density function is unknown. One way is to approximate the densities by means of histograms, but an arbitrary histogram would not be the best way, because it can cause underestimation or overestimation of the empirical mutual information.

The mutual information defined in Eq.(C.2.2) takes a value between 0 and infinity, $0 \leqslant I(X,Y) \leqslant +\infty$, which makes the comparisons difficult between different samples. In this context, [45, 47, 48, 59] among others, defined and used a standard measure for the mutual information:

$$\lambda = \Big(1 - exp[-2I(X,Y)]\Big)^{\frac{1}{2}}. \tag{C.2.3}$$

Note that λ captures the overall dependence, both linear and non-linear, between X and Y. This measure varies between 0 and 1 being thus directly comparable to the linear correlation coefficient, ρ, based on the relationship between the measures of information theory and variance analysis. According to the properties of the mutual information, and because independence is one of the most valuable concepts in econometrics, we can construct an independence test based on the following hypothesis:

$$\begin{cases} H_0: & I(X,Y) = 0 \\ H_1: & I(X,Y) > 0 \end{cases} \tag{C.2.4}$$

If $P(x,y) = P(x)P(y)$, then H_0 is not rejected and the independence between the variables is found. Otherwise, if $P(x,y) \neq P(x)P(y)$, then H_1 is accepted and we reject the null hypothesis of independence.

Another technique to check whether there are autocorrelations in

time series is based on the investigation of the fractal structure in time series and is related to the scaling exponent H, called Hurst exponent, and sometimes denoted as α. In the following we consider two methods that can be used to calculate α.

C.3 Detrended fluctuation analysis

Detrended fluctuation analysis (DFA) gives a measure of the time-dependent fluctuations in a series [50]. In fact, the DFA is a scaling analysis method used to quantify long-range power-law correlations in signal embedded in a nonstationary time series [134]. In the past few years, the DFA has been used as a method of correlation analysis to uncover long range power-law correlations in financial time series [135]–[140].

The idea of the DFA was first proposed to investigate the long-range dependence in coding and non-coding DNA nucleotide sequences [50]. The advantages of DFA over many methods are that it permits the detection of the long-range correlations embedded in seemingly non-stationary time series, and also avoids the spurious detection of apparent long-range correlations that are an artifact of non-stationarity. The method employed to derive the DFA was carried out through the following procedure.

Consider a time series $Y_T = (y_1, \ldots, y_T)$ of length T. As the DFA has been originally designed for the DNA walk, one needs to consider a related random walk series. Therefore, the series Y_T is first integrated

after subtracting the average value $\bar{y} = \frac{1}{T}\sum_{i=1}^{T} y_i$:

$$x_i = \sum_{j=1}^{i}(y_j - \bar{y}) \qquad (i = 1, \ldots, T) \tag{C.3.1}$$

Next, the integrated series $X_T = (x_1, \ldots, x_T)$ is divided into subseries (boxes) of equal length n. A polynomial function $\hat{x}_{i_n}$, which represents the local trend in each box, is fitted to the series x_i. Linear, quadratic, cubic, or higher order functions can be used in the fitting procedure. Next, the integrated series x_i is deterended by subtracting the local trend $\hat{x}_{i_n}$ in each box; $x_i - \hat{x}_{i_n}$. The DFA analysis is a modified root-mean-square (RMS) analysis of a random walk. The RMS fluctuation of the integrated and detrended time series is calculated by

$$F_n = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(x_i - \hat{x}_{i_n})^2}. \tag{C.3.2}$$

This computation is repeated over all time scales (box sizes) to characterize the relationship between F_n, the average fluctuation, as a function of box size n. Typically, F_n will increase with box size n. A linear relationship on a log-log plot indicates the presence of power law (fractal) scaling. Under such conditions, the fluctuations can be characterized by a scaling exponent α, the selfsimilarity parameter, which is the slope of the line relating log F_n to log n [?]. A power-law relation between F_n and the box size n represents the presence of scaling:

$$F_n \sim n^{\alpha}. \tag{C.3.3}$$

Equation (C.3.3) enables to calculate α exponent directly from log-

log linear fit; $\log F_n \sim \alpha \log n$. The value of α indicates the degree of the correlation in the series: If $\alpha = 0.5$, there is no correlation and the signal is uncorrelated (white noise); if $0 < \alpha < 0.5$, the series is anticorrelated (antipersistence); if $0.5 < \alpha < 1$, indicates positive long-range power-law correlations (persistence) and $\alpha = 1.5$ for the Brownian walk (for more information see, for example, [15, 16, 38]).

C.4 Detrended Moving Average Method

The Detrended Moving Average (DMA) method [51,52] is a relatively new method that is widely used to quantify correlation in a non-stationary economic time series with underlying trends in the series [141,142]. The DMA method determines whether data follow the trend, and how deviations from the trend are correlated. The first step of the DMA method is to detect trends in data using a moving average. There are two kinds of moving average procedure; simple moving average and weighted moving average. Here we use the backward and the simple moving average. The simple backward moving average, for a window of size n, is

$$m_{i_n} = \frac{1}{n} \sum_{j=1}^{n-1} x_{i-j} \,, \tag{C.4.1}$$

where x_i is the integrated series defined in Eq.(C.3.1). In fact, the m_{i_n} at each data point i depends only on the past $n-1$ values of the series. In the next step we detrend the series by subtracting the trend m_{i_n} from the integrated series x_i; $x_i - m_{i_n}$. Therefore, the rms fluctuation of the integrated and detrended time series is calculated by

$$F_n = \sqrt{\frac{1}{T-n+1}\sum_{i=n}^{T}(x_i - m_{i_n})^2}. \qquad \text{(C.4.2)}$$

Again, repeating the calculation for different n, we obtain the fluctuation function F_n. A power law relation between the fluctuation function F_n and the scale n indicates a self-similar behavior. The DMA technique looks very similar to the DFA. The main difference one meets here is that instead of linear or polynomial detrendisation procedure in equally sized boxes, one uses moving average of a given length n. Unlike the DFA analysis, the DMA method is used without any assumptions on the type of trends, the probability distribution, or other features of the series.

Appendix D

APPLICATION OF SSA FOR THE FABRICATED METAL SERIES IN GERMANY

We shall now use the Fabricated metal series for Germany as an example to illustrate the selection of the SSA parameters and to show the reconstruction of the original series in detail. To perform the analysis, we have used the SSA software[1]. Fig. D.1 presents the series, indicating a complex trend and strong seasonality.

Selection of the window length L

The window length L is the only parameter in the decomposition stage. Knowing that the time series may have a periodic component with an integer period, to achieve a better separability of this periodic component it is advisable to take the window length proportional to that period. For example, the assumption that there is an annual periodicity in the series suggests that we must pay attention to the frequencies $k/12$ ($k = 1, ..., 12$). As it is advisable to choose L reasonably large (but smaller than $T/2$ which is 162 in this case), we choose $L = 120$.

[1] http://www.gistatgroup.com/cat/index.html

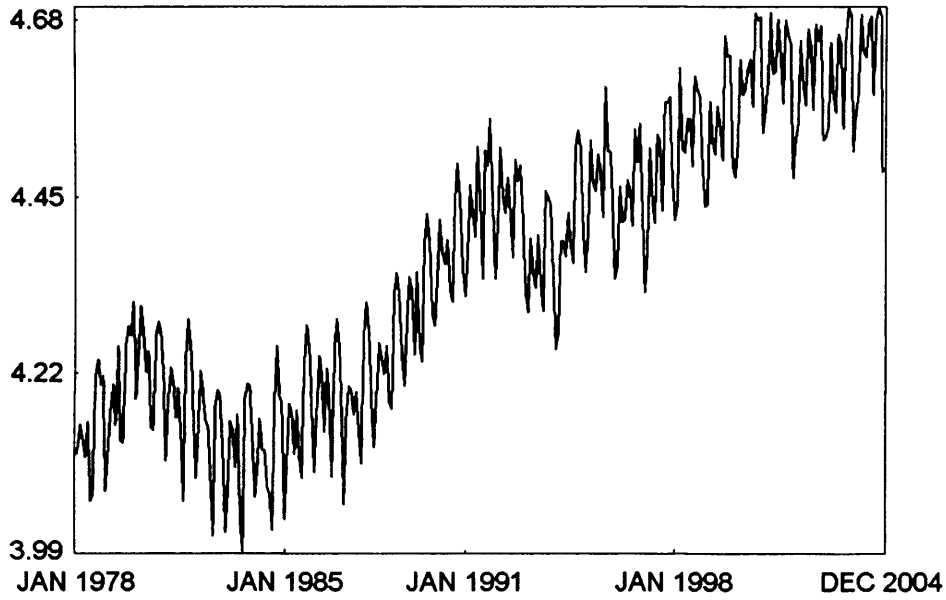


Figure D.1. Fabricated metal series in Germany

Selection of r

Auxiliary information can be used to choose the parameters L and r. Below we briefly explain some methods that can be useful in the separation of the signal from noise. Usually a harmonic component produces two eigentriples with close singular values (except for the frequency 0.5 which provides one eigentriple with the saw-tooth singular vector). Another useful insight is provided by checking breaks in the eigenvalue spectra. Additionally, a pure noise series typically produces a slowly decreasing sequence of singular values.

Choosing $L = 120$ and performing SVD of the trajectory matrix $\mathbf{X}$, we obtain 120 eigentriples, ordered by their contribution (share) in the decomposition. Fig. D.2 depicts the plot of the logarithms of the 120 singular values.

Here a significant drop in values occurs around component 19 which could be interpreted as the start of the noise floor. Six evident pairs,

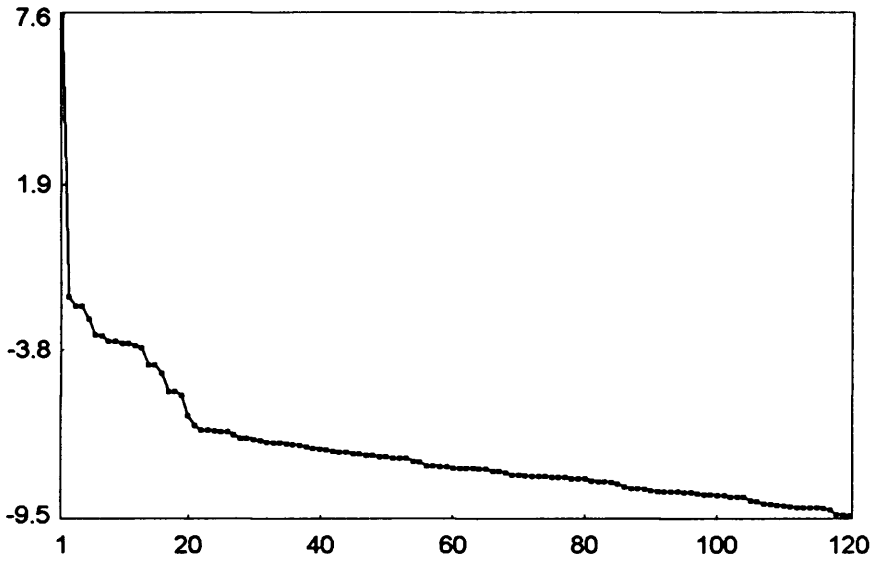


Figure D.2. Logarithms of the 120 eigenvalues.

with almost equal leading singular values, correspond to six (almost) harmonic components of the series: eigentriple pairs 3-4, 6-7, 8-9, 10-11, 14-15 and 17-18 are related to the harmonics with specific periods (we show later that they correspond to the periods of 6, 4, 12, 3, 36 and 2.4 months).

Another way of grouping is to examine the matrix of the absolute values of the w-correlations. Fig. D.3 shows the w-correlations for the 120 reconstructed components in a 20-grade grey scale from white to black corresponding to the absolute values of correlations from 0 to 1. Based on this information, we select the first 18 eigentriples for the reconstruction of the original series and consider the rest as noise.

The principal components (shown as time series) of the first 18 eigentriples are shown in Fig. D.4. Consider a pure harmonic with a frequency w, certain phase, amplitude and the ideal situation where the period $P = 1/w$ is a divisor of both the window length L and $K = T - L + 1$. In this ideal situation, the left eigenvectors and

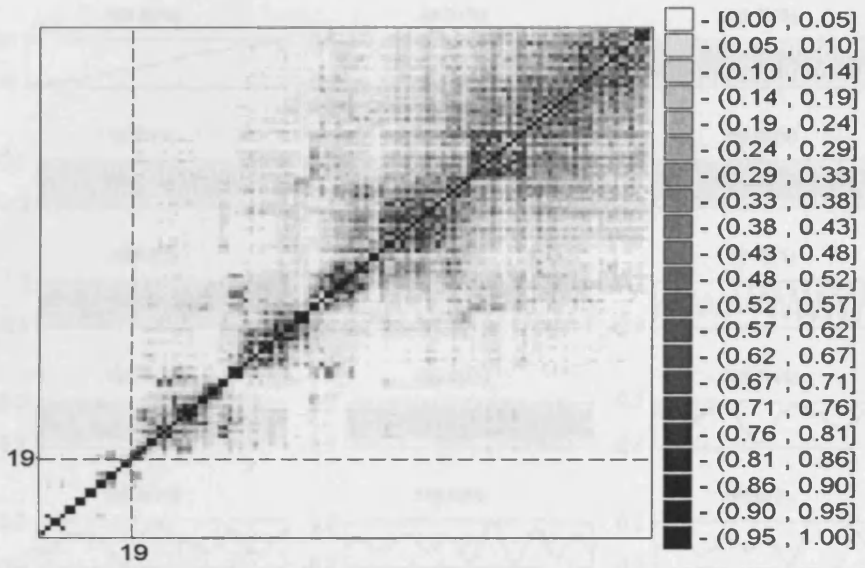


Figure D.3. Matrix of w-correlations for the 120 reconstructed components.

principal components have the form of sine and cosine sequences with the same period P and the same phase. Thus, the identification of the components that are generated by a harmonic is reduced to the determination of these pairs.

Fig. D.5 depicts the scatterplots of the paired principal components in the series, corresponding to the harmonics with periods 6, 4, 12, 3, 36 and 2.4 months. They are ordered by their contribution (share) in the SVD step (from left to right).

The periodograms of the paired eigentriples (3-4 , 6-7, 8-9, 10-11 and 17-18) also confirm that the eigentriples correspond to the periods of 6, 4, 12, 3, 36 and 2.4 months.

Identification of trend, harmonics and noise components

Trend is a slowly varying component of a time series which does not contain oscillatory components. Henceto capture the trend in the series,

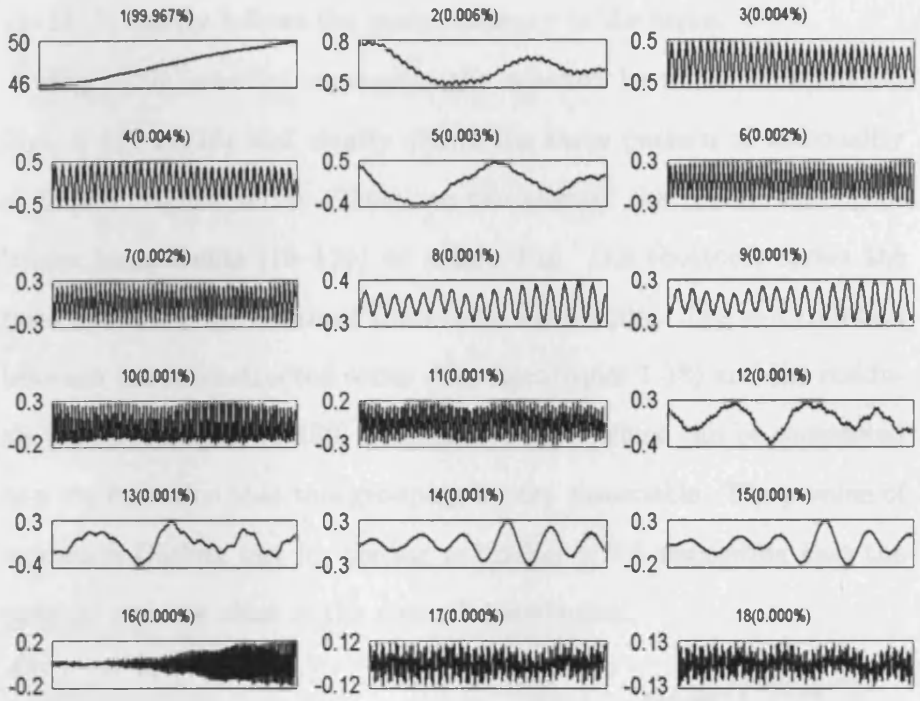


Figure D.4. The first 18 principal components plotted as time series

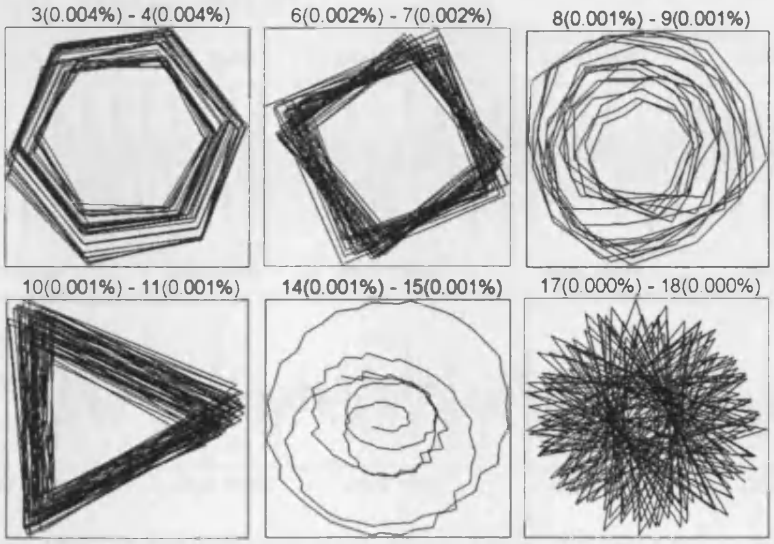


Figure D.5. Scatterplots (with lines connecting consecutive points) corresponding to the paired harmonic principal components.

we should look for slowly varying eigenvectors. Fig. D.6 (top) shows the extracted trend which is obtained from the eigentriples 1, 2, 5, and

12–13. It clearly follows the main tendency in the series.

Fig. D.6 (middle) represents the selected harmonic components (3,4, 6–11, 14–18) and clearly shows the same pattern of seasonality as in the original series. Thus, we can classify the rest of the eigentriples components (19–120) as noise. Fig. D.6 (bottom) shows the residuals which are obtained from these eigentriples. The w-correlation between the reconstructed series (the eigentriples 1-18) and the residuals (the eigentriples 19-120) is equal to 0.0006, which can be considered as a confirmation that this grouping is very reasonable. The p-value of Anderson-Darling test for testing normality is 0.6 suggesting that the residual series is close to the normal distribution.

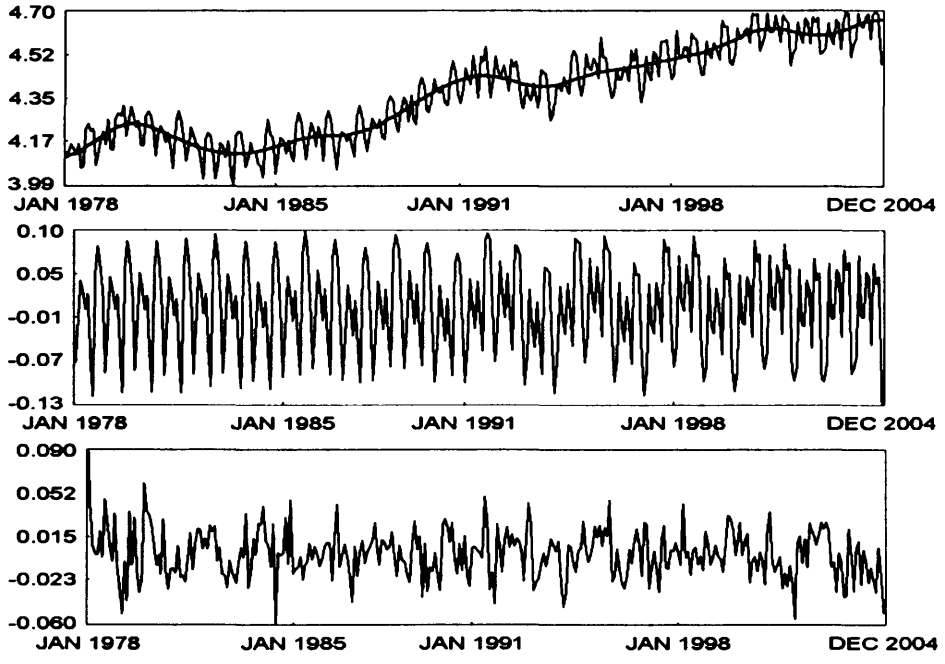


Figure D.6. Reconstructed trend (top), harmonic (middle) and noise (bottom).

Appendix E

INDUSTRIAL PRODUCTION SERIES

The two-digit categories examined in this research are given in the following table. For more information about these series and some graphs depicting them (up to 1995), see [66].

Short name	Detail
Food product (da15)	Manufacture of food products and beverages
Chemicals (dg24)	Manufacture of chemicals and chemical product
Basic metals (dj27)	Manufacture of basic metals
Fabricated metal (dj28)	Manufacture of fabricated metal products
Machinery (dk29)	Manufacture of machinery and equipment N.E.C.
Electrical machinery (dl31)	Manufacture of electrical machinery and apparatus N.E.C.
Vehicles (dm34)	Manufacture of motor vehicles, trailers and semi-trailers
Electricity and gas (e40)	Electricity, gas and water supply

Table E.1. Industrial production series.

Appendix F

SEPARABILITY

F.0.1 Weak and strong separability

Let $Y_T^{(1)}$ and $Y_T^{(2)}$ be time series of length T and $Y_T = Y_T^{(1)} + Y_T^{(1)}$. Under the choice of window length L, each of the series $Y_T^{(1)}$, $Y_T^{(2)}$ and F_N generates an L-trajectory matrix: $\mathbf{X^{(1)}}$, $\mathbf{X^{(2)}}$ and $\mathbf{X}$.

Denote by $\mathcal{L}^{(L,1)}$ and $\mathcal{L}^{(L,2)}$ the linear spaces spanned by the columns of the trajectory matrices $\mathbf{X^{(1)}}$ and $\mathbf{X^{(2)}}$. Similar notation $\mathcal{L}^{(K,1)}$ and $\mathcal{L}^{(K,2)}$ will be used for the spaces spanned by the columns of the transposed matrices $(\mathbf{X^{(1)}})'$ and $(\mathbf{X^{(2)}})'$, $K = N - L + 1$.

If $\mathcal{L}^{(L,1)} \perp \mathcal{L}^{(L,2)}$ and $\mathcal{L}^{(K,1)} \perp \mathcal{L}^{(K,2)}$, then we say that the series $Y_T^{(1)}$ and $Y_T^{(2)}$ are *weakly L-separable.*

For brevity, we shall use the term 'separability' instead of 'weak L-separability' in cases when no ambiguity occur.

Let us elucidate the last definition. Suppose that the series $Y_T^{(1)}$ and $Y_T^{(2)}$ are L-separable. Consider certain SVDs of the trajectory matrices $\mathbf{X^{(1)}}$ and $\mathbf{X^{(2)}}$:

$$\mathbf{X^{(1)}} = \sum_k \sqrt{\lambda_{1k}}\, U_{1k} V_{1k}', \quad \mathbf{X^{(2)}} = \sum_k \sqrt{\lambda_{2k}}\, U_{2k} V_{2k}'. \qquad \text{(F.0.1)}$$

Then

$$\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)} = \sum_k \sqrt{\lambda_{1k}}\, U_{1k} V_{1k}' + \sum_m \sqrt{\lambda_{2m}}\, U_{2m} V_{2m}'. \quad \text{(F.0.2)}$$

Therefore, we can conclude that (F.0.2) is an SVD of the matrix $\mathbf{X}$. Thus, the representation $Y_N = Y_N^{(1)} + Y_N^{(2)}$ is natural from the viewpoint of the SVD of the matrix $\mathbf{X}$.

If $Y_T^{(1)}$ and $Y_T^{(2)}$ are weakly L-separable and $\lambda_{1k} \neq \lambda_{2m}$ for all k and m, then we say that $Y_T^{(1)}$ and $Y_T^{(2)}$ are *strongly L-separable.* The difference between separability and strong separability can be expressed as follows. If separability occurs, then an SVD of the matrix $\mathbf{X}$ exists such that we can group its terms in a proper way and obtain $Y_T^{(1)}$ and $Y_T^{(2)}$ in terms of their trajectory matrices $\mathbf{X^{(1)}}$ and $\mathbf{X^{(2)}}$. In the case of strong separability, we can obtain $Y_T^{(1)}$ and $Y_T^{(2)}$ for any SVD of the trajectory matrix $\mathbf{X}$. In this section we study features of weak separability. Suppose that nonzero series $Y_T^{(1)}$ and $Y_T^{(2)}$ are weakly L-separable. Denote by d_1, d_2 the ranks of the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. Since $d_1 + d_2 = \text{rank } \mathbf{X} \leq L$ both d_1 and d_2 do not exceed $L - 1$. Therefore, the time series $Y_T^{(1)}$ and $Y_T^{(2)}$ have L-ranks smaller than L.

Let $K = N - L + 1$. Time series $Y_T^{(1)}$ and $Y_T^{(2)}$ are weakly L-separable if and only if

1. for any $0 \leq k, m < K - 1$

$$y_k^{(1)} y_m^{(2)} = y_{k+L}^{(1)} y_{m+L}^{(2)}; \quad \text{(F.0.3)}$$

2. for any $0 \le m \le K-1$

$$y_m^{(1)} y_0^{(2)} + \ldots + y_{m+L-1}^{(1)} y_{L-1}^{(2)} = 0; \qquad \text{(F.0.4)}$$

3. for any $0 \le k, m < L-1$

$$y_k^{(1)} y_m^{(2)} = y_{k+K}^{(1)} y_{m+K}^{(2)}; \qquad \text{(F.0.5)}$$

4. for any $0 \le m \le L-1$

$$y_m^{(1)} y_0^{(2)} + \ldots + y_{m+K-1}^{(1)} y_{K-1}^{(2)} = 0. \qquad \text{(F.0.6)}$$

Proof.

By definition, weak L-separability is equivalent to the matrix equalities

$$(\mathbf{X^{(1)}})' \mathbf{X^{(2)}} = \mathbf{0_{KK}} \qquad \text{and} \qquad \mathbf{X^{(1)}} (\mathbf{X^{(2)}})' = \mathbf{0_{LL}}. \qquad \text{(F.0.7)}$$

Taking the first equality in (F.0.7) we obtain the condition

$$y_k^{(1)} y_m^{(2)} + \ldots + y_{k+L-1}^{(1)} y_{m+L-1}^{(2)} = 0, \quad 0 \le k, m \le K-1, \qquad \text{(F.0.8)}$$

which is equivalent to (F.0.3), (F.0.4). The second equality in (F.0.7) is equivalent to (F.0.5), (F.0.6).

Let us now turn back to our problem. For a fixed length L, consider a certain SVD of the noisy series Y_T of length T, and assume that the series Y_T is a sum of two series S_T and N_T; $Y_T = S_T + N_T$. In this case, separability of the series S_T and N_T means that we can split the matrix terms of the SVD of the trajectory matrix $\mathbf{X}$ into two different groups, so that the sums of terms within the groups give the trajectory

matrices $\mathbf{S}$ and $\mathbf{N}$ of the series S_T and N_T, respectively.

As we mentioned above the separability immediately implies that each row of the trajectory matrix $\mathbf{S}$ of the first series is orthogonal to each row of the trajectory matrix $\mathbf{N}$ of the second series, and the same holds for the columns. Since rows and columns of trajectory matrices are subseries of the corresponding series, the orthogonality condition for the rows (and columns) of the trajectory matrices $\mathbf{S}$ and $\mathbf{N}$ is just the condition of orthogonality of any subseries of length L (and $K = T - L + 1$) of the series S_T to any subseries of the same length of the series N_T (the subseries of the time series must be considered here as vectors).

If this orthogonality holds, then we shall say that the series S_T and N_T are *weakly separable*. If all the singular values of the trajectory matrix $\mathbf{X}$ are different, then the conditions for weak separability and strong separability coincide. Below, for brevity, we shall use the term 'separability' for 'weak separability'.

Strong separability of two series S_T and N_T is equivalent to the fulfillment of the following two conditions: (a) the series S_T and N_T are weakly separable, and (b) the collections of the singular values of the trajectory matrices $\mathbf{S}$ and $\mathbf{N}$ are disjoint.

In practice, the lack of strong separability (under the presence of the weak separability, perhaps, approximate) becomes essential when the matrix $\mathbf{XX}'$ has two close eigenvalues. This leads to an instability of the SVD computations.

F.0.2 Approximate and asymptotic separability

Exact separability does not happen for real-life series and in practice we can only assume approximate separability. Next we consider the characteristics that reflect the degree of separability.

For a fixed window length L, the definition of weak separability of series $Y_T^{(1)}$ and $Y_T^{(2)}$ is formulated in terms of orthogonality for their subseries. This leads to the natural concept of *approximate separability* of two time series. For any series $Y_T = (y_1, \ldots, y_T)$ we set

$$Y_{i,j} = (y_i, \ldots, y_j), \quad 1 \le i \le j < T. \tag{F.0.9}$$

Let $Y_T^{(1)} = (y_1^{(1)}, \ldots, y_T^{(1)})$, $Y_Y^{(2)} = (y_1^{(2)}, \ldots, y_T^{(2)})$. For $i, j \ge 1$ and $M \le T - \max(i, j)$ we set

$$\rho_{i,j}^{(M)} = \frac{\left(Y_{i,i+M-1}^{(1)}, Y_{j,j+M-1}^{(2)}\right)}{||Y_{i,i+M-1}^{(1)}||\,||Y_{j,j+M-1}^{(2)}||} \tag{F.0.10}$$

under the assumption that the denominator is positive.

The notation $(\cdot, \cdot)$ stands for the usual inner product of Euclidean vectors and $||\cdot||$ is the Euclidean norm. If the denominator in (F.0.10) is equal to zero, then we assume that $\rho_{i,j}^{(M)} = 0$.

The number $\rho_{i,j}^{(M)}$ has the sense of the cosine of the angle between the vectors $Y_{i,i+M-1}^{(1)}$ and $Y_{j,j+M-1}^{(2)}$. Using the statistical terminology, we can call $\rho_{i,j}^{(M)}$ the *correlation coefficient* between $Y_{i,i+M-1}^{(1)}$ and $Y_{j,j+M-1}^{(2)}$.

Time series $Y_T^{(1)}, Y_T^{(2)}$ are (weakly) *ϵ-separable for the window length* L if

$$\rho^{(L,K)} \stackrel{\text{def}}{=} \max\left(\max_{1 \le i,j \le K} |\rho_{i,j}^{(L)})|, \max_{1 \le i,j \le L} |\rho_{i,j}^{(K)})|\right) < \epsilon. \tag{F.0.11}$$

If the number ϵ is small, then the series are *approximately separable.* Of course, if separable time series $Y_T^{(1)}$ and $Y_T^{(2)}$ are slightly perturbed, they become ϵ-separable with some small ϵ. Suppose that the parameters L and T provide weak separability of the series $Y_T^{(1)}, Y_T^{(2)}$. Then another way from separability to approximate separability is in a small perturbation of the parameters L and T.

The concept of approximate separability has its asymptotic variant. Consider infinite time series $Y^{(1)} = (y_1^{(1)}, \ldots, y_t^{(1)}, \ldots)$ and $Y^{(2)} = (y_1^{(2)}, \ldots, y_t^{(2)}, \ldots)$. For each $T > 2$ let the series $Y_T^{(1)}$ and $Y_T^{(2)}$ consist of the first N terms of the series $Y^{(1)}$ and $Y^{(2)}$, respectively. Choosing a sequence of window lengths $1 < L = L(T) < T$, we obtain the related sequence of the *maximum correlation coefficients* $\rho_T = \rho^{(L,K)}$ defined by (F.0.11).

If there exists a sequence $L = L(T)$ such that $\rho_T \to 0$ as $T \to \infty$, then the time series $Y^{(1)}$ and $Y^{(2)}$ are called *asymptotically separable.* If $Y^{(1)}$ and $Y^{(2)}$ are asymptotically separable for any choice of L such that $L \to \infty$ and $K \to \infty$, then they are called *regularly asymptotically separable.* Conditions for regular asymptotic separability can be written as follows; when $T_1, T_2 \to \infty$, then

$$\rho(T_1, T_2) \stackrel{\text{def}}{=} \max_{i,j<N_1} \frac{\left| \sum_{k=0}^{T_2-1} y_{i+k}^{(1)} y_{j+k}^{(2)} \right|}{\sqrt{\sum_{k=0}^{T_2-1} \left(y_{i+k}^{(1)}\right)^2} \sqrt{\sum_{k=0}^{T_2-1} \left(y_{j+k}^{(2)}\right)^2}} \to 0. \quad \text{(F.0.12)}$$

In the case of exact separability, the orthogonality of rows and columns of the trajectory matrices $\mathbf{S}$ and $\mathbf{N}$ means that all pairwise inner products of their rows and columns are zero. In statistical language,

this means that the noncentral covariances (and therefore, noncentral correlations — the cosines of the angles between the corresponding vectors) are all zero. This implies that we can consider as a characteristic of separability of two series $\mathbf{s_t}$ and $\mathbf{n_t}$ the *maximum correlation coefficient* $\rho^{(L,K)}$, that is the maximum of the absolute value of the correlations between the rows and between the columns of the trajectory matrices of these two series (as usual, $K = N - L + 1$).

We shall say that two series $\mathbf{s_t}$ and $\mathbf{n_t}$ are *approximately separable* if all the correlations between the rows and the columns of the trajectory matrices $\mathbf{S}$ and $\mathbf{N}$ are close to zero.

Bibliography

[1] Plaut, G., and Vautard, R. (1994). Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere, *J. Atmos. Sci.*, **51**, pp. 210-236.

[2] Box, G. E. P., and Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*, Holden-Day.

[3] Brockwell, P. J., and Davis R. A. (2002). *Introduction to Time Series and Forecasting*, 2nd edition. Springer.

[4] Cao, L. Y., and Soofi, A. (1999). Nonlinear deterministic forecasting of daily dollar exchange rates, *International Journal of Forecasting*, **15**(4), pp. 421–430.

[5] Soofi, A., and Cao, L. Y. (2002). Nonlinear Forecasting of Noisy Financial Data, in Soofi and Cao (eds.), *Modeling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Kluwer Academic Publishers, Boston.

[6] Hsieh, D. A. (1991). Chaos and nonlinear Dynamics: Application to Financial Markets, *Journal of Finance*, **46**, pp. 1839–1877.

[7] Scheinkman, J., and LeBaron, B. (1989). Nonlinear Dynamics and Stock Returns, *Journal of Business*, **62**, pp. 311–337.

[8] Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, New York - London.

[9] Broomhead, D. S., and King, G. (1986). Extracting qualitative dynamics from experimental data, *Physica* D, **20**, pp. 217–236.

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

[11] Elsner, J. B., and Tsonis, A. A. (1996). *Singular Spectrum Analysis, A New Tool in Time Series Analysis*, Plenum Press, New York and London.

[12] Hassani, H., and Zhigljavsky, A. (2009). Singular Spectrum Analysis: Methodology and Application to Economics Data, *Journal of System Science and Complexity*, **22**(3), pp. 372–394.

[13] Moskvina, V. G., and Zhigljavsky, A. (2003). An algorithm based on singular spectrum analysis for change-point detection, *Communication in Statistics - Simulation and Computation*, **32**(4), pp. 319–352.

[14] Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison, *Journal of Data Science*, **5**(2), pp. 239–257.

[15] Hassani, H. (2010). Singular Spectrum Analysis Based on the Minimum Variance Estimator, *Nonlinear Analysis: Real World Applications*, Forthcoming.

[16] Hassani, H., Heravi, S., and Zhigljavsky, A. (2009). Forecasting

European Industrial Production with Singular Spectrum Analysis, *Internation Journal of Forecasting*, **25**(1), pp. 103–118.

[17] Alexandrov, Th., and Golyandina, N. (2004). The automatic extraction of time series trend and periodical components with the help of the Caterpillar-SSA approach. *Exponenta Pro 3-4* (In Russian.), pp. 54–61.

[18] www.gistatgroup.com

[19] Marple-Jr, S. L. (1987). *Digital Spectral Analysis*, Prentice Hall, New Jersey.

[20] Bouvet, M., and Clergeot, H. (1988). Eigen and singular value decomposition technique for the solution of harmonic retrieval problems. In E. F. Deprettere (Ed.), *SVD and Signal processing: Algorithm, Applications and Architectures*, North-Holland, Amsterdam, pp. 93–114.

[21] Madisetti, V. K., and Lloyd, E. (Eds.) (1998). *The Digital Signal Processing Handbook*, CRC Press, Boca Raton.

[22] Brillinger, D. (1975). *Time Series. Data Analysis and Theory*, Holt, Rinehart and Winston, Inc., New York.

[23] Subba Rao, T. (1976). Canonical factor analysis and stationary time series models. *Sankhya: The Indian Journal of Statistics*, **38**B, pp. 256–271.

[24] Subba Rao, T., and Gaber, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models.* Speringer-Verlag.

[25] Priestly, M. B. (1991). *Spectral Analysis and Time Series.* Academic Press, London.

[26] Wei, W. W. S. (1990). *Time Series Analysis: Univaiate and Multivariate Methods.* Addision -Wessly, New York.

[27] Cutler, C. D., and Kaplan (Eds.). (1997). *Nonlinear Dynamics and Time Series: Building a Bridge between the natural and Statistical Science.* American Statistical Society, Providence, Rhode Island.

[28] Abarbanel, H. D. I. (1996). *Analysis of observed Chaotic Data.* Springer, New York.

[29] Tong, H. (1993). *Nonlinear Time Series Analysis: A Dynamical System Approach.* Oxford University Press, Oxford.

[30] Kantz, H., and Schreiber, T. (1997). *Nonlinear Time Series Analysis.* Cambridge University Press, Cambridge.

[31] Hassani, H., Dionisio, A., and Ghodsi, M. (2009). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, *Nonlinear Analysis: Real World Applications*, doi:10.1016/j.nonrwa.2009.01.004.

[32] Alonsoa, F. J., Del Castilloa, J. M., and Pintado, P. (2004). Application of singular spectrum analysis to the smoothing of raw kinematic signals. *Journal of Biomechanics*, **38**, pp. 1085–1092.

[33] Ghodsi, M., Hassani, H., Sanei, S., and Hick, Y. (2009). The use of noise information for detecting temporomandibular disorder, *Biomedical Signal Processing and Control*, **4**, pp. 79–85.

[34] Hassani, H., Zokaei, M., von Rosen, D., Amiri, S., and Ghodsi, M. (2009). Does Noise Reduction Matter for Curve Fitting in Growth

Curve Models?, *Computer Methods and Program in Biomedicine*, **96**(3), pp. 173–181.

[35] Brock, W. A., Dechert, W. D., and Scheinkman, J. (1987). A Test for Independence Based on the Correlation Dimension. Department of Economics, University of Wisconsin, University of Houston and University of Chicago, 1987, (Revised Version, 1991: Brock, W.A., W.D. Dechert, J. Scheinkman and B. LeBaron).

[36] Abhyankar, A., Copeland, L. S., and Wong, W. (1995). Nonlinear Dynamics in Real-Time Equity Market Indices: Evidence From the United Kingdom, *The Economic Journal*, **105**, pp. 864–880.

[37] Cecen, A. A., and Erkal, C. (1996). Distinguishing between stochastic and deterministic behavior in foreign exchange rate returns: Further evidence, *Economics Letters*, **51**, pp. 323–329.

[38] Bouchaud, J. P., and Potters, M. (2001). More Stylized Facts of Financial Markets: Leverage Effect and Downside Correlations, *Physica A*, **299**, pp. 60–70.

[39] Drozdz, S., Grümmer, F., Ruf, F., and Speth, J. (2001). *Proceedings of the Empirical Science of Financial Fluctuations*, Tokyo ,Springer Verlag.

[40] Mantegna, R., and Stanley, E. (2000). *Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge Univ. Press, Cambridge, UK.

[41] Ullah, A. (2002). Uses of Entropy and Divergence Measures for Evaluating Econometric Approximations and Inference, *Journal of Econometrics*, **107**, pp. 313–326.

[42] McCauley, J. (2003). Thermodynamic Analogies in Economics and Finance: Instability of Markets, *Physica A*, **329**, pp. 199–212.

[43] McCauley, J. (2004). *Dynamics of Markets: Econophysics and Finance*, Cambridge Univ. Press, Cambridge.

[44] Maasoumi, E., and Racine, J. (2002). Entropy and Predictability of Stock Market Returns, *Journal of Econometrics*, **107**, pp. 291–312.

[45] Granger, C. W. J., and Lin, J. (1994). Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models, *Journal of Time Series Analysis*, **15**, pp. 371–384.

[46] Urbach, R. (2000). *Footprints of Chaos in the Markets—Analysing Non-linear Time Series in Financial Markets and Other Real Systems* ,Prentice-Ha ll, London, 2000.

[47] Darbellay, G., and Wuertz, D. (2000). The Entropy as a Tool for Analysing Statistical Dependence's in Financial Time Series, *Physica A* , **287**, pp. 429–439.

[48] Dionisio, A., Menezes, R., and Mendes, D. A. (2004). Mutual Information: a Measure of Dependency for Nonlinear Time Series, *Physica A*, **344**, pp. 326–329.

[49] Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Long range correlations in DNA sequences, *Phys. Rev. E49*, **2**, pp. 1685–1689.

[50] Peng, C. K., Havlin, S., Stanley, H. E., and Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, *Chaos*, **5**, pp. 82–87.

[51] Alessio, E., Carbone, A., Castelli, G., and Frappietro, V. (2002). Scaling Properties of Long-Range Correlated Noisy Signals, *Eur. Phys. Jour. B*, **27**, pp. 197–200.

[52] Carbone, A., Castelli, G., and Stanley, H. E. (2004). Analysis of clusters formed by the moving average of a long-range correlated time series, *Phys. Rev. E*, **69**, pp. 0261051-4 .

[53] Soofi, A., and Cao, L. (eds.) (2002). *Modelling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Kluwer Academic Publishers: Boston.

[54] Yan, J. (2005). Asymmetry, Fat-tail, and Autoregressive Conditional Density in Financial Return Data with Systems of Frequency Curves. Tech. Rep. 355, Department of Statistics and Actuarial Science, University of Iowa, http://www.stat.uiowa.edu/techrep/tr355.pdf.

[55] Pagan, A. (1996). The econometrics of financial markets, *Journal of Empirical Finance*, **3**, pp. 15–102.

[56] Fama, E. (1970). Stock Returns, Expected Returns and Real Activity, *Journal of Finance*, **25**, pp. 383–417.

[57] Scheinkman, J., and LeBaron B. (1989). Nonlinear Dynamics and Stock Returns, *Journal of Business*, **62**, pp. 311–337.

[58] Hsieh, D. A. (1991). Chaos and Nonlinear Dynamics: Application to the Financial Markets, *Journal of Finance*, **46**, pp. 1839–1877.

[59] Dionisio, A., Menezes, R., and Mendes, D. A. (2006). Entropy-Based Independence Test, *Nonlinear Dynamics*, **44**, pp. 351–357.

[60] Hénon, M. (1976). A two-dimensional mapping with a strange attractor, *Communications in Mathematical Physics*, **50**, pp. 69–77.

[61] Yang, S. R., and Brorsen, B. W. (1993). Nonlinear dynamics of daily futures prices: Conditional heteroskedasticity or chaos?, *The Journal of Futures Markets*, **13**, pp. 175–191.

[62] Grech, D., Mazur, Z. (2009). Comparison study of DFA and DMA methods in analysis of autocorrelations in time series, arXiv:cond-mat/0507395v1.

[63] Sozanski, M., and Zebrowski, J. (2005). On the Application of DFA to the Analysis of Unimodal Maps, *Acta Physica Polonica B*, **36**(5), pp. 1803–1822.

[64] Szpiro, G.(1997). Noise in unspecified, non-linear time series, *Journal of Econometrics*, **78**, pp. 229–255.

[65] Davis, R. A., and Mikosch, T. (2000). The Sample Autocorrelations of Financial Time Series Models, Nonlinear and Nonstationary Signal Processing, W.J. Fitzgerald, R.L. Smith, A.T. Walden, P. Young, editors, Cambridge University Press, Cambridge, England, pp. 247–274.

[66] Osborn, D. R., Heravi, S., and Birchenhall, C. R. (1999). Seasonal unit roots and forecasts of two-digit European industrial production. *International Journal of Forecasting*, **15**, pp. 27–47.

[67] Heravi, S., Osborn, D. R., and Birchenhall, C. R. (2004). Linear Versus Neural Network Forecasts for European Industrial Production Serie,. *International Journal of Forecasting*, **20**, pp. 435–446.

[68] Scott, M. J., and Goldsmith, O. S. (1987). Assessing Reginal Econometric Model: A Discussion and Application, Annals of Regional Sceince, **21**(1), pp. 1–21.

[69] Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997). Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting*, **13**, pp. 281–291.

[70] Harvey, D. I., Leybourne, S. J., and Newbold, P. (1998). Tests for Forecast Encompassing, *Journal of Business and Economic Statistics*, **16**, pp. 254-59.

[71] Krane, S. D. (2003) An evaluation of real GDP forecasts. http://www.chicagofed.org/publications/economicperspectives/2003/1qeppart1.pdf.

[72] Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: do they fit out-of-sample? *Journal of International Economics*, **14**, pp. 3–24.

[73] Frankel, J. and K. A. Froot. (1990) Chartists, Fundamentalists, and Trading in the Foreign Exchange market, *American Economic Review*, **80**, pp. 181–185.

[74] Frankel, J. A. (1979). On the mark: a theory of floating exchange rates based on real interest differentials, *American Economic review*, **69**, pp. 610–622.

[75] Corbae, D. and S. Ouliaris (1988). Cointegration and tests of purchasing power parity, *Review of Economics and Statistics*, **70**, pp. 508–511.

[76] Chinn, M. D. (1997). Paper pushers or paper money? Empirical assessment of fiscal and monetary models of exchange rate determination. *Journal of Policy Modeling*, **19**, pp. 51–78

[77] Soofi, A. and Cao, L. (2002a). (eds.) *Modelling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Kluwer Academic Publishers: Boston.

[78] Soofi, A., and Galka, A. (2003) Measuring the Complexity of Currency Markets by Fractal Dimension Analysis, *International Journal of Theoretical and Applied Finance*, **6**(6), pp. 553–563.

[79] Cheung, Y., M. D. Chinn, and Pasual, A. G. (2005). Empirical exchange rate models of nineties: Are any fit to survive?, *Journal of International Money and Finance*, **24**, pp. 1150–1175.

[80] Soofi, A., and Cao, L. (2002b). Prediction and Volatility of Black Market Currencies: Evidence from Renminbi and Rial Exchange Rates. *International Journal of Theoretical and Applied Finance*, **5**, pp. 659–666.

[81] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37**(3), pp. 424-438.

[82] Phillips, A. W. (1958). The relationship between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957, *Economics*, **25**, pp. 283–99.

[83] Samuelson, P., and Solow, R. (1960). Analytical aspects of anti-inflation policy, *American Economic Review*, **50**, pp. 177–94.

[84] Phelps, E. (1967). Phillips curve, expectations of inflation, and optimal inflation over time, *Economica*, **135**, pp. 254–281.

[85] Friedman, M. (1968). The Role of Monetary Policy, *American Economic Review*, **58**, pp. 1–17.

[86] Stock, J., and Watson, M. W. (1999). Forecasting inflation, *Journal of Monetary Economics*, **44**, pp. 293–335.

[87] Gordon, R. J. (1997). The time-varying NAIRU and its implications for economic policy, *Journal of Economic Perspectives*, **11**, pp. 11–32.

[88] Atkeson, A., and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation?, Federal Reserve Bank of Minneapolis, *Quarterly Review*, **25**(1), pp. 2–11.

[89] Rudd, J. and Whelan, K. (2007) Modeling inflation dynamics: A critical review of recent research, *Journal of Money, Credit, and Banking*, **39**:155-170.

[90] Fair, R. C. (2008). Testing price equations, *European Economic Review*, **52**, pp. 1424–1437.

[91] Stock, J., and Watson, M. W. (2005). Has inflation become harder to forecast? paper presented in Conference on *Quantitative Evidence on Price Determination*, Board of Governors of the Federal Reserve System, September 29–30, Washington DC.

[92] Gavin, W. T., and Kliesen, K. L. (2006). Forecasting inflation and output: Comparing data-rich models with simple rules, *Research Division, Federal Reserve Bank of St. Luis, Working Paper Series, 2006–054B.*

[93] Thomas, L. B. (1999). Survey measures of expected U.S. inflation. *Journal of Economic Perspectives*, **13**, pp. 125-144.

[94] Mehra, Y. P.(2002). Survey measures of expected inflation: revisiting the issues of predictive content and rationality. Federal Reserve Bank of Richmond Economic Quarterly 88, 1736.

[95] Ang, A., Bekaert, G., and Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics* **54**, pp. 1163-1212

[96] Kim, C., and Nelson, C. R. (1999). Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle, *The Review of Economics and Statistics*,**81**, pp. 608-616.

[97] Blanchard, O. J., and Simon, J. A. (2001). The Long and Large Decline in U.S. Output Volatility, Brookings Papers on Economic Activity, **1**, pp. 135–164.

[98] Cogley, T. W., and Sargent, T. (2005). Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War II U.S., *Review of Economic Dynamics*, **8**, pp. 262–302.

[99] Tufts, D. W., Kumaresan, R., and Kirsteins, I. (1982). Data adaptive signal estimation by singular value decomposition of a data matrix, *Proceedings of the IEEE*, **70**(6), pp. 684-685.

[100] Cadzow, J. A. (1988). Signal enhancementa composite property mapping algorithm, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36**(1), pp. 49-62.

[101] De Moor, B. (1993). The singular value decomposition and long and short spaces on noisy matrices, *IEEE Transaction on Signal Processing*, **41**(9), pp. 2826–2838.

[102] Van Huffel, S. (1993). Enhanced resolution based on minimum variance estimation and exponential data modeling, *Signal Processing*, **33**(3), pp. 333-355.

[103] Jensen, S. H., Hansen, P. C., Hansen, S. D. and Sørensen, J. A. (1995). Reduction of Broad-Band Noise in Speech by Truncated QSVD, *IEEE Transactions on Speech and Audio Processing, November*, **6**, pp. 439–448.

[104] Stewart, G. W., and Sun, J. (1990). *Matrix Perturbation Theory.* San Diego, CA: Academic.

[105] Golub, G., and van Loan, C. (1996). *Matrix computations*, third edition, The Johns Hopkins University Press, London.

[106] Ash, J. C. K., Smyth, D. J., and Heravi, S. (1997). The accuracy of OECD forecasts for Japan. *Pacific Economic Review*, **2**(1), pp. 25–44.

[107] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, (2nd ed), Springer.

[108] Chu, T., Danks, D. and Glymour, C. (2005). Data Driven Methods for Nonlinear Granger Causality: Climate Teleconnection Mechanisms, Technical report (CMU-PHIL-171).

[109] Clements, M. P., and Smith, J. (1999). A Monte Carlo investigation of forecasting performance of emprical SETAR model. *Journal of Applied Econometrics*, **14**, pp. 123–141.

[110] Diks, C. and DeGoede, J. (2001). A general nonparametric bootstrap test for Granger causality. In: HW broer, B Krauskopf, G Vegter (Eds.), Global Analysis of Dynamical Systems (Institute of Physics Publishing, Bristol, UK), pp. 391–403.

[111] Diks, C. and Panchenko, V. (2005). A note on the Hiemstra-Jones test for Granger non-causality. *Studies in Nonlinear Dynamics and Econometrics*, **9** , art. 4.

[112] Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing, *Journal of Economic Dynamics and Control*, **30**, pp. 1647-1669.

[113] Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd ed.). New York: Wiley.

[114] Granger, C. W. J. (1980). Testing for causality: A personal viewpoint, *Journal of Economic Dynamics and Control*, **2**, pp. 329–352.

[115] Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, **49**(5), pp. 1639-1664.

[116] Su, L., and White, H. (2008). A Nonparametric Hellinger Metric Test For Conditional Independence, *Econometric Theory*, **24**(4), pp. 829–864

[117] Hassani, H., Soofi, A., and Zhigljavsky, A. (2009). Predicting Daily Exchange Rate with Singular Spectrum Analysis. *Nonlinear Analysis: Real World Applications*, doi:10.1016/j.nonrwa.2009.05.008

[118] Patterson, K. D. (1992). Revisions to the components of the trade balance for the United Kingdom, *Oxford Bulletin of Economics and Statistics*, **54**, pp. 103–120.

[119] Patterson, K. D. (1994). Consumers' expenditure on nondurables and services and housing equity withdrawal in the United Kingdom, *Manchester School of Economic and Social Studies*, **62**. pp. 251–274.

[120] Patterson, K. D. (1995a). An integrated model of the data measurement and data generation processes with an application to consumers' expenditure, *Economic Journal*, **105**, pp. 54–76.

[121] Patterson, K. D. (1995b). A State Space Approach to Forecasting the Final Vinatege of Revised Data with an Application to the Index of Industrial Production, *Journal of Forecasting*, **14**, pp. 337–350.

[122] Patterson, K. D. (1995c). Forecasting the final vintage of real personal disposable income: A state space approach, *International journal of forecating*, **11**, pp. 395–405.

[123] Patterson, K. D., and Heravi, , S. M. (1991a). Data revisions and the expenditure components of GDP, *Economic Journal*, **101**, pp. 887–901.

[124] Patterson, K. D., and Heravi, S. M. (1991b). Are different vintages of data on the components of GDP cointegrated?, *Economics Letters*, **35**, pp. 409–413.

[125] Patterson, K. D., and Heravi, S. M. (1992). Efficient forecasts or measurement errors? Some evidence for revisions to the United Kingdom GDP growth rates, Manchester School of Economic and Social Studies, **60**, pp. 249–263.

[126] Su, L. and White, H (2008). Nonparametric Hellinger metric test for conditional independence, *Econometric Theory*, **24**, pp. 829-864.

[127] Toda, H. Y., and Phillips, P. C. B. (1991). Vector Autoregressions and Causality: A Theo- retical Overview and Simulation Study, *Econometric Reviews*, **13**, pp. 259–285.

[128] 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**, pp. 35-62.

[129] Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica*, **50**(4), pp. 987–1007.

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

[131] Bollerslev, T., Chou Kenneth, R. Y. and Kroner, F. (1992). ARCH modeling in finance : A review of the theory and empirical evidence, Journal of. *Econometrics*, **52**, pp. 5–59.

[132] Fan, J. and Yao, Q. (2005). *Non linear time series : nonparametric and parametric methods*, Springer, NY.

[133] Tavares, A. B., Dias Curto, J., and Tavares, G. N. (2008). Modelling Heavy Tails and Asymmetry Using ARCH-Type Models with Stable Paretian Distributions, *Nonlinear Dynamics*, **51**, pp. 231–243.

[134] Havlin, S., Amaral, L. A. N., Ashkenazy, Y., Goldberger, A. L., Ivanov, P. Ch., Peng, C. K., and Stanley, H. E. (1999). Application of statistical physics to heartbeat diagnosis, *Physica A*, **274**, pp. 99.

[135] Liu, Y., Gopikrishnan, P., Cizeau, P., Meyer, M., Peng, C. K., and Stanley, H. E. (1999). Scaling of the distribution of fluctuations of financial market indices, *Phys. Rev. E*, **60**, pp. 1390.

[136] Janosi, I. M., Janecsko, B., and Kondor, I. (1999). Statistical analysis of 5 s index data of the Budapest Stock Exchange, *Physica A*, **269**, pp. 111.

[137] Ausloos, M., Vandewalle, N., Boveroux, P., Minguet, A. and Ivanova, K. (1999). Application of Statistical Physics to Economic and Financial Topics, *Physica A*, **274**, pp. 229.

[138] Roberto, M., Scalas, E., Cuniberti, G., and Riani, M. (1999). Effect of nonstationarities on detrended fluctuation analysis, *Physica A*, **269**, pp. 148.

[139] Vandewalle, N., Ausloos, M., and Boveroux, P. (1999). The moving averages demystiÿed, *Physica A*, **269**, pp. 170.

[140] Ausloos, M., and Ivanova, K. (2000). Introducing False EUR and False EUR exchange rates, *Physica A*, **286**, pp. 353.

[141] Kwon, K., and Kish, R. J. (2002). Technical Trading Strategies and Return Predictability: NYSE, *Applied Financial Economics*, **12**, pp. 639–653.

[142] Yeh, A. B., Lin, D. K. J., Zhou, H., and Renkataramani, C. (2003). Quantifying signals with power-law correlations: A comparative study of detrended fluctuation analysis and detrended moving average techniques, *J. App. Stat*, **30**, pp. 507.