

BOURNEMOUTH UNIVERSITY

Bayesian Singular Spectrum Analysis with State Dependent Models

by

Donya Rahmani

Supervisors

Dr. Damien Fay

Dr. Paul D.Yoo

A thesis submitted in partial fulfilment for the
degree of Doctor of Philosophy

in the

Faculty of Science and Engineering
Department of Computing and Informatics

December, 2016

BOURNEMOUTH UNIVERSITY

ABSTRACT

FACULTY OF SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTING AND INFORMATICS

Doctor of Philosophy

by Donya Rahmani

The analysis of time series using Singular Spectrum Analysis has become an important area of statistics with application in a variety of fields such as economics, geophysics, engineering, medicine and many others. In fact, in this method there is no need to make any statistical assumptions such as the stationarity of the series or normality of the residuals. Therefore, SSA is recognised as an extremely practical tool which can be used to solve problems without considering any parametric model. At its core SSA depends on an eigenvector decomposition of the covariance matrix of a time series which may be utilised for forecasting via a linear recurrent formula. However, many time series exhibit structural breaks which interfere with a linear continuation of the time series although the underlying data generating process may not have changed. In addition, in a multivariate setting there is the added complication of combining time series. In this case the linear recurrence relationships of each time series may either reinforce each other or alternatively may lead to degraded forecasts.

In this thesis a state dependent model is proposed under the assumption that if a system moves from one homogeneous state to another rapidly that this transition may be tracked using a Bayesian model in which the state transitions are state dependent. In addition, it is proven that for basic SSA the linear recurrent coefficients are biased and that this bias decays linearly with the samples. Empirically, the state dependent model shows far superior performance over two multivariate data sets.

In the second part of the thesis component matching is examined. The core issue is how to identify which time series to group together without testing every possible combination. Geographical information resulted in superior forecasts on USA unemployment time series via a spatial SSA model. Subsequent research into data driven methods to group the time series concludes that a novel variant of the self organising map leads to a significant improvement over methods based on standard techniques like tensor analysis and joint diagonalisation.

Contents

Nomenclature	xiii
Glossary	xv
Declaration and Copyright	xvii
Acknowledgements	xix
1 Introduction	1
1.1 Motivation for research	4
1.2 Contributions	4
1.3 Thesis structure	5
1.4 List of publications and conference papers	6
2 Background	9
2.1 Singular Spectrum Analysis	9
2.1.1 Decomposition	11
2.1.1.1 Embedding	12
2.1.1.2 Projection pursuit	14
Singular Value Decomposition	15
Principal component analysis	16
Independent component analysis	17
Joint Diagonalisation	18
Tucker Decomposition	19
2.1.2 Reconstruction	22
2.1.2.1 Grouping	22
2.1.2.2 Diagonal Averaging	22
2.1.3 Parameter Selection	23
2.1.3.1 Embedding dimension	23
Separability (w-correlation)	24
2.1.3.2 Selection of components for reconstruction	25
2.1.4 Linear Recurrent Formula (LRF)	27
2.1.5 Extensions to basic SSA	28
2.1.5.1 Vector SSA	28
2.1.5.2 Bootstrap SSA	29
2.1.5.3 Monte Carlo SSA	30
2.1.5.4 Spatial SSA	31
2.2 Bayesian Forecasting	32

2.2.1	Linear vs Nonlinear	33
2.2.2	Nonparametric vs parametric	35
2.2.3	State Space Models	36
2.2.3.1	Kalman Filtering	38
2.2.3.2	State Dependent Models	40
2.2.4	Bayesian SSA	43
3	Preliminary analysis	45
3.1	Introduction	45
3.2	Bootstrap LRF	46
3.3	Empirical Results	47
3.3.1	Synthetic time series	47
3.3.2	Real data sets	50
3.4	Conclusion	53
4	Bayesian Singular Spectrum Analysis with a State Dependent model	55
4.1	Introduction	55
4.2	The Multivariate Bayesian SSA algorithm	56
4.2.1	SSA model	56
4.2.2	Linear Recurrent Formula in the presence of a structural break	58
4.2.3	State dependent format of Linear Recurrent Formula parameters	61
4.3	Empirical results; Univariate	63
4.3.1	Synthetic data	63
4.3.2	Real data	66
4.3.2.1	Data, descriptive statistics and structural Breaks	66
4.3.2.2	Forecasting results	72
	BSSA vs Boot SSA and SSA	72
	BSSA vs ARIMA, ARFIMA, ETS and GARCH	73
	Empirical cumulative distribution function	73
4.3.3	Summary and conclusion on univariate case	73
4.4	Empirical results; Multivariate	74
4.4.1	Synthetic data	74
4.4.2	Real data	75
4.4.2.1	Variance Inflation Factor	76
4.4.2.2	Engle-Granger cointegration test	83
4.4.2.3	Forecasting results	84
	MSSA-2 vs MBSSA-2	84
	VAR vs VECM	84
	MSSA-8 vs MBSSA-8	85
	ICA-MSSA-8 vs ICA-MBSSA-8	85
	Empirical cumulative distribution function	85
4.4.3	Summary and conclusion	85
5	Latent and Reduced Space Multivariate Singular Spectrum Analysis	89
5.1	Introduction	89
5.2	Spatial weights	90
5.2.1	Data, descriptive statistics and graphs	91

	Spatial autocorrelation	93
	5.2.1.1 Empirical Results	95
5.3	Projection pursuit using Joint Diagonalisation	96
	5.3.1 Empirical Results	98
	5.3.1.1 JD versus SVD	98
	5.3.1.2 Clustering states on δ_i	99
	Gaussian mixture model	99
	Kmeans clustering	101
	Hierarchical clustering	102
	5.3.1.3 Forecasting clustered states	103
5.4	Projection pursuit with Tensor decomposition	104
	5.4.1 Tensor Analysis	104
	5.4.2 Empirical results	105
	5.4.2.1 Clustering with Tensor-modes	106
	5.4.2.2 Forecasting	106
5.5	Projection pursuit with a self-organising map	107
	5.5.1 Self Organising Eigenspace Map	107
	5.5.1.1 SOFM vs SOEM	108
	5.5.2 Empirical results	110
	5.5.2.1 Clustering with SOEM	110
	5.5.2.2 Forecasting clustered states	111
	5.5.3 Summary and conclusion	111
6	Conclusion	117
A		121
	A.1 Granger-Newbold test	121
B		125
	Bibliography	147

List of Figures

1.1	A time series and its delayed caterpillar plot.	2
1.2	SSA modes and coefficients.	2
1.3	A forecast of the demonstration time series using SSA.	3
3.1	Realisation of the synthetic data.	48
3.2	Three time series used for comparison.	50
3.3	Cloud map of ϕ_1 for the Vehicle time series.	52
4.1	Realisation of synthetic data.	64
4.2	Coefficients of the LRF for univariate synthetic data	65
4.3	Effect of noise level on SSA and BSSA for the synthetic data.	65
4.4	Empirical cumulative distribution functions of absolute values of univariate forecast errors	78
4.5	Coefficients of the LRF for bivariate synthetic data	79
4.6	Coefficients of the LRF for multivariate synthetic data	80
4.7	Empirical cumulative distribution functions of absolute values of multivariate forecast errors	86
5.1	92
5.2	Box-plot of United States Unemployment Rate.	93
5.3	Moran Scatterplot for United States Unemployment Rate.	94
5.4	Distribution of cosine distances between the strongest eigenvector of every state	97
5.5	Distribution of δ_i (kernel smoothing is employed for the overall average.) .	100
5.6	States clustered with GMM.	101
5.7	States clustered with Kmeans.	101
5.8	Dendogram.	102
5.9	States Hierarchically clustered.	103
5.10	Tucker transform.	105
5.11	States clustered with Tensor-Kmeans.	106
5.12	self-organising eigenspace map.	108
5.13	Distribution of winning nodes over several iterations.	113
5.14	Distance measure, $\langle \{U_{i,j}\}, C \rangle$, for four sample American states.	114
5.15	Average distance to s_2	114
5.16	States clustered with SOEM.	115
A.1	Industrial Production Indicators for France.	122
A.2	Industrial Production Indicators for Germany.	123
A.3	Industrial Production Indicators for the UK.	124

B.1 Unemployment rates for United States. 126

List of Tables

3.1	Post-sample forecast accuracy measures for synthetic exponential time series.	49
3.2	Post-sample forecast accuracy measures for the real time series.	52
4.1	Descriptive statistics of Industrial Production Indicators for France. . . .	66
4.2	Descriptive statistics of Industrial Production Indicators for Germany. . .	67
4.3	Descriptive statistics of Industrial Production Indicators for the UK. . . .	67
4.4	Univariate post-sample forecast accuracy measures for France.	69
4.5	Univariate post-sample forecast accuracy measures for Germany.	70
4.6	Univariate post-sample forecast accuracy measures for the UK.	71
4.7	Summary statistics for univariate out-of-sample forecasting accuracy measures for Industrial Production Indicators for all three countries.	74
4.8	Variance Inflation Factor to test multi-collinearity.	76
4.9	P-values of Augmented Dickey-Fuller test.	76
4.10	Engle-Granger cointegration test.	77
4.11	Multivariate post-sample forecast accuracy measures for France.	81
4.12	Multivariate post-sample forecast accuracy measures for Germany.	82
4.13	Multivariate post-sample forecast accuracy measures for the UK.	83
5.1	Summary statistics for out-of-sample forecasting accuracy measures for unemployment rate series.	96
5.2	Summary statistics (JD, Tensor and SOEM clusters vs SVD).	99
5.3	Summary statistics for clustering based on JD.	103
5.4	Summary statistics comparing spatial SSA and tensor SSA using a single eigenspace/cluster.	107
5.5	Summary statistics (MSSA, MBSSA, and BD vs SOEM).	111
B.1	Descriptive statistics for Unemployment Rate.	127
B.2	Local Moran test for Unemployment Rate.	128
B.3	Bivariate post-sample forecast accuracy measures with spatial weights for UR of states AL,AR,AZ,CA,CO,CT,DE,FL.	129
B.4	Bivariate post-sample forecast accuracy measures with spatial weights for UR of states GA,IA,ID,IL,IN,KS,KY,LA.	130
B.5	Bivariate post-sample forecast accuracy measures with spatial weights for UR of states MA,MD,ME,MI,MN,MO,MS,MT.	131
B.6	Bivariate post-sample forecast accuracy measures with spatial weights for UR of states NC,ND,NE,NH,NJ,NM,NV,NY.	132
B.7	Bivariate post-sample forecast accuracy measures with spatial weights for UR of states OH,OK,OR,PA,RI,SC,SD,TN.	133

B.8	Bivariate post-sample forecast accuracy measures with spatial weights for UR of states TX,UT,VA,VT,WA,WI,WV,WY.	134
B.9	Multivariate post-sample forecast accuracy for clustered states with JD for UR of states AL,AR,AZ,CA,CO,CT,DE,FL.	135
B.10	Multivariate post-sample forecast accuracy for clustered states with JD for UR of states GA,IA,ID,IL,IN,KS,KY,LA.	136
B.11	Multivariate post-sample forecast accuracy for clustered states with JD for UR of states MA,MD,ME,MI,MN,MO,MS,MT.	137
B.12	Multivariate post-sample forecast accuracy for clustered states with JD for UR of states NC,ND,NE,NH,NJ,NM,NV,NY.	138
B.13	Multivariate post-sample forecast accuracy for clustered states with JD for UR of states OH,OK,OR,PA,RI,SC,SD,TN.	139
B.14	Multivariate post-sample forecast accuracy for clustered states with JD for UR of states TX,UT,VA,VT,WA,WI,WV,WY.	140
B.15	Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states AL,AR,AZ,CA,CO,CT,DE,FL.	141
B.16	Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states GA,IA,ID,IL,IN,KS,KY,LA.	142
B.17	Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states MA,MD,ME,MI,MN,MO,MS,MT.	143
B.18	Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states NC,ND,NE,NH,NJ,NM,NV,NY.	144
B.19	Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states OH,OK,OR,PA,RI,SC,SD,TN.	145
B.20	Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states TX,UT,VA,VT,WA,WI,WV,WY.	146

Nomenclature

ϕ_i	Coefficients of LRF
λ	Eigenvalue
\mathcal{T}	Vector of samples prior to structural break
\mathcal{Q}	Amplitudes of the structural breaks
C	$\mathbf{X}\mathbf{X}^T$
K	$N-L+1$
N	Length of time series
\mathbf{X}_i	$[y_i, \dots, y_{i+L}]^T$
\mathbf{X}	$[\mathbf{X}_1, \dots, \mathbf{X}_K]$
M	Number of time series in multivariate system
U	Eigenvector
$J(y)$	Negative entropy
\hat{Q}_i	Cumulant matrices
L	Window length or embedding dimension
r	Number of selected eigencomponents
Φ	$(\phi_{L-1}, \phi_{L-2}, \dots, \phi_1)$
$U_i^{\nabla L}$	$(u_{i,1}, \dots, u_{i,L-1})$
π_i	last component of U_i
$\mathcal{P}^{(v)}$	linear operator
F_t	Transition matrix
F_θ	Jacobian matrix of F
S_N	Signal
E_N	Noise
\mathbf{Y}	$Y_N^{(m)} = (y_1^{(m)}, \dots, y_N^{(m)})$
\mathbf{J}	Matrix of ones
h	Number of forecast horizons

Glossary

AR	Autoregressive
ARCH	Autoregressive conditional heterogeneity
ARFIMA	Autoregressive fractional integrated moving average
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
ASR	Asymmetric ratio
B.M.	Basic metals
BD	Border Distance
BootSSA	Bootstrap Singular Spectrum Analysis
BSSA	Bayesian Singular Spectrum Analysis
CANDECOM	Canonical decomposition
CPC	Common principal component
E.M.	Electrical machinery
E&G	Electricity and gas
EEG	Electroencephalogram
EKF	Extended Kalman filtering
ETS	Exponential state smoothing
F.M	Fabricated metals
F.P.	Food products
FI	Forecast interval
GARCH	Generalised autoregressive conditional heterogeneity
GDP	Gross domestic products
GMM	Gaussian mixture model
GN	Granger-Newbold
ICA	Independent component analysis
ID	Inverse distance
IPI	Industrial production indicator
JADE	Joint approximate diagonalisation
JD	Joint diagonalisation
KF	Kalman filtering
L_{FI}	Lower bound of forecast interval
LRF	Linear recurrent formula

MBSSA	Multivariate Bayesian Singular Spectrum Analysis
MS-MSSA	Multi-state Multivariate Singular Spectrum Analysis
MSSA	Multivariate Singular Spectrum Analysis
MTM	Multi-taper method
PAR	Periodic autoregressive
PARAFAC	parallel decomposition
PCA	Principal component analysis
PSSA	Posterior Singular Spectrum Analysis
RMSE	Root mean square error
RRMSE	Ratio of root mean square error
SD	Standard deviation
SDM	State dependent model
SNR	Signal to noise ratio
SOFM	Self-organising feature map
SOME	Self-organising eigenspace map
SSA	Singular Spectrum Analysis
SSD	Singular Spectrum decomposition
STAR	smooth transition autoregressive
SVD	Singular value decomposition
TVAR	Threshed vector autoregressive
U_{FI}	Upper bound of forecast interval
UKF	Unscented Kalman filtering
UR	Unemployment rate
VAR	Vector autoregressive
VECM	Vector error correction model
VIF	Variance inflation
VSSA	Vector singular spectrum analysis

Declaration and Copyright

The work in this thesis has been undertaken by myself and has not been accepted nor concurrently submitted in candidature for any other award. Conference and journal publications related to this work are referenced. This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Acknowledgements

First, I want to express my sincere thank to my supervisor, Dr. Damien Fay for all his support during my PhD study, his patience, immense knowledge and valuable guidance. I could not have imagined having a better supervisor for my PhD. His door was always open to me whenever I had a question or problem about my research or writing. I place on record, my sincere thank to my examiners, Prof. Simon Wilson and Dr. Emili Balaguer-Ballester for their valuable comments and guidance.

Next, I must express my very profound gratitude to my parents and my brothers, Damoon and Danesh, for providing me with unfailing support and continuous encouragement throughout my years of study and through my life. This accomplishment would not have been possible without them. Thank you very much.

I also take this opportunity to express my gratitude to all of the department of computing, faculty of science and engineering for their help and support. I am grateful to Santander for its funding for my joint project with Cardiff University, Prof. Saeed Heravi. I am very thankful to him for sharing time and expertise during my work with him. I would like to thank to all my friends and colleagues, especially Barras Stone, Peter Howard-Jones, Azade Pazouki, Sevil Yesiloglu, Navid Aslani, Mastoureh Fathi and all others.

Last but not least, I will be ever grateful to my beloved grandma and I am sad that she cannot see me graduate. Her memory will be with me always.

*To My lovely parents:
Ghasem Rahmani & Zohre Akbari
for their unceasing support and love*

Chapter 1

Introduction

SSA was first introduced by (Broomhead & King, 1986). Since then the analysis of time series using Singular Spectrum Analysis has become an important area of statistics with application in a variety of fields such as economics, geophysics, engineering, medicine and many others. In fact, in this method there is no need to make any statistical assumptions such as the stationarity of the series or normality of the residuals. Therefore, SSA is recognised as an extremely practical tool which can be used to solve problems without considering any parametric model. SSA can also be used for smoothing, finding trends of diverse resolution and simultaneous extraction of harmonics and trend components, forecasting and so on (Golyandina et al., 2001).

The SSA method is based on decomposing a time series into three types of components: trend, harmonics and noise. The method then reconstructs the original series and forecasts based on the reconstructed series. One advantage of SSA is the ability to reconstruct multiple series which have shared dynamics and which may have complex seasonal components and/or trends. This characteristic is suitable to retrieve the underlying sources in the analysis of, for example, electroencephalogram (EEG) signals (Kouchaki et al., 2015; Teixeira et al., 2005), or to capture major periodicity change of the El Niño/Southern Oscillation (ENSO) in weather forecasting to mention but a few applications.

As a brief introduction to the method consider the time series, Y , shown in Figure 1.1(a) below. A delayed version of this time series is shown in Figure 1.1(b) and this resembles a caterpillar (an alternate name for SSA). If we consider the delayed versions of Y as a multivariate system and calculate their singular value decomposition, this then allows us to approximate the original series via an eigenvector reconstruction. The first three eigenvector reconstruction terms, called *modes*, are shown in Figure 1.2(a). Note how these decompose this series into its trend (the strongest eigenvector) and two cyclical components. In this particular example, we use 24 delayed copies of the signal and these may be related to the current value via a set of coefficients (details given in Chapter 2),

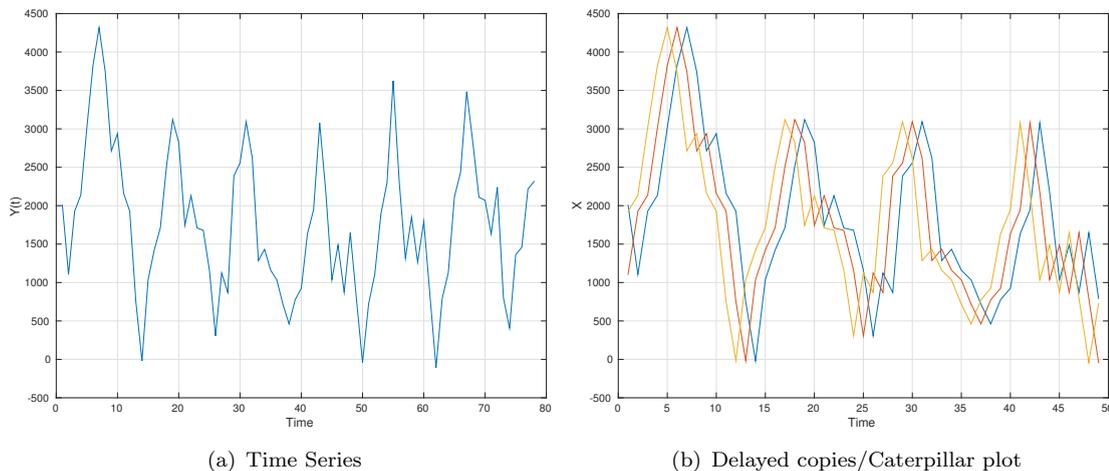


FIGURE 1.1: A time series and its delayed caterpillar plot.

ϕ_Y which are shown in Figure 1.2(b). Note that ϕ_Y are data dependent and will change as new data arrives (this model differs significantly from a standard AR model) also note that we have made no assumptions about stationarity or the distribution of the residual. As ϕ_Y relates the past values of the series to the present it may be used recursively to

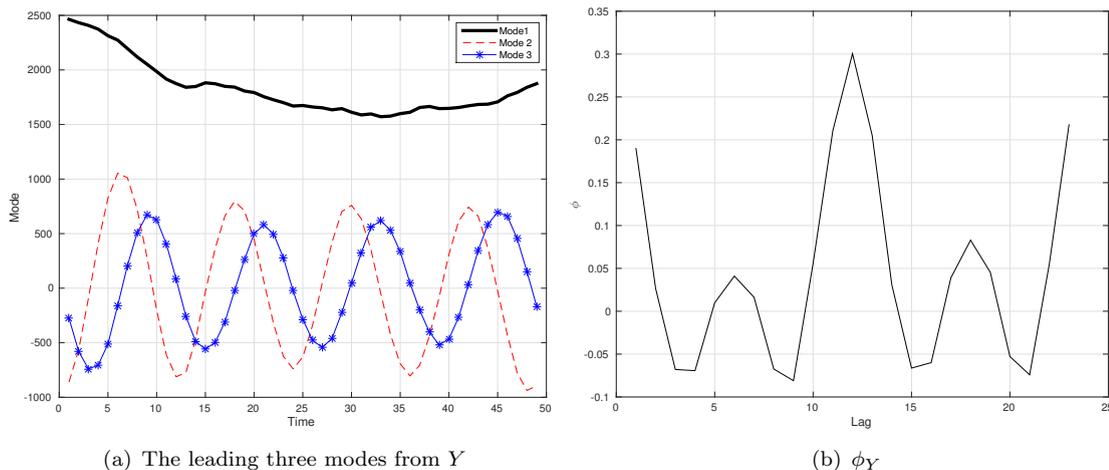


FIGURE 1.2: SSA modes and coefficients.

produce a forecast. One such forecast is shown in Figure 1.3. This whole process is known as the basic SSA forecasting algorithm, full details of which are provided in Chapter 2.

For multivariate time series, an important challenge in using SSA is to recognise which different time series should be modelled together. The key issue is whether the modes (or components) in those different time series reinforce each other or whether they effectively add noise to the estimated (common) modes. The issue here is whether components are *matched*. In SSA this can be determined specifically by clustering of covariance matrices, a topic which we explore. This issue goes to the heart of *separability* (See Section 2.1.3.1), in which modes which are distinct but similar between time series may be inseparable.

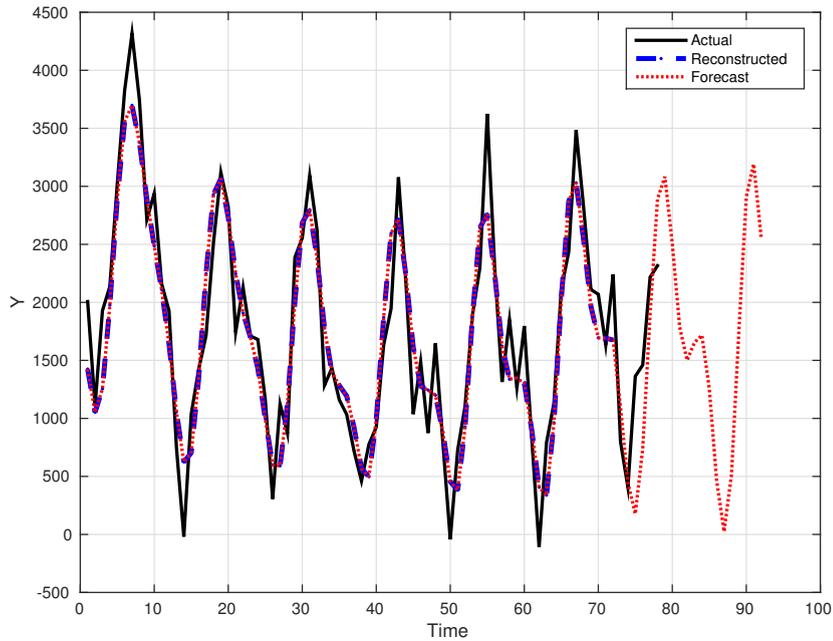


FIGURE 1.3: A forecast of the demonstration time series using SSA.

Another issue concerns forecasting time series with special structure, such as turning point, abrupt periodicity shifts and structural breaks (Golyandina et al., 2001). In such a case the modes, their associated eigenvectors and the SSA forecasts no longer match the future evolution of the system. SSA basically assumes that a time series evolves according to an underlying state space system. The major effect of a shock is that a system will be transferred from one region of the state space to another region over a short period of time; in essence it entails not a fixed state space system but one which sequentially adapts to the transition in the region of the state space. A state space model that jumps from one region of the state space to another is called a state-dependent model (SDM). The SDM technique essentially uses a Bayesian approach to update the parameter estimation of state dependents models and in this work we examine the applicability of such a model to SSA.

Common to both issues raised above, multivariate grouping and special structures, is differences in the modes either in time or across series. However, this focus on modes has not been fully explored and is the core of this thesis.

Two different kinds of data sets will be used to evaluate the efficiency of the model: Industrial production Indicators for France, Germany and the UK; which allows comparison of the results over the structural break which occurs due to the recession in 2008. The other time series used in this study is the unemployment rates of the USA states, obtained by the Federal Reserve Bank of St. Louis. These series also exhibit structural breaks, but in addition, they are distributed geographically favourably (in the sense that they resemble a grid allowing a full spatial comparison). It is of interest to determine

the significance of various geographical characteristics in matching modes as compared with approaches based on the data itself.

1.1 Motivation for research

As mentioned earlier SSA can be used for smoothing, trend extraction etc. Here the focus is on forecasting alone. By extrapolating the information from past and present, forecasts can help policy-makers in dealing with uncertain future conditions.

Structural breaks are a common occurrence in many time series and occur when a sudden change hit the system. The structural break can be accommodated by allowing the parameters to change rapidly at the time of the event as well as allowing a change in their state. Such models essentially provides much better forecasts, because they better approximate the underlying structure of the given series under study. Hence, an interesting direction for our research is to explicitly explore Bayesian forecasting applied to MSSA/SSA in the presence of a structural break.

The issue of matched components has been discussed in SSA research before but has not been fully explored. In a multivariate time series situation such as that used in Chapter 5 with 48 time series there are for example $48(2^{47} - 1) = 6.7 \times 10^{15}$ combinations which would have to be tested to determine the best grouping for each series with respect to its partners. Thus a more computationally efficient approach is required for estimating which modes may or may not match.

1.2 Contributions

The main contributions of this thesis are:

1. The extension of Bootstrapping to the coefficients in SSA (Bootstrapping LRF) is explored and found empirically to give a minor improvement in terms of forecasting.
2. Theorem 4.1 is proposed which shows that SSA is biased in the presence of a structural break and that this bias decays linearly. Specifically, the deviation between the first eigenvalue of C_X and \check{C}_X is $\lambda_1 - \check{\lambda}_1 \approx \mathcal{TL}(Q^2 - 2Q\bar{y})$. Furthermore, this deviation decreases as $\mathcal{O}(k/N)$.
3. It is shown empirically that using an State Dependent Model (SDM) produces superior forecasts than SSA in the presence of a structural break alleviating the issues raised with SSA in 4.1. SDM assumes that the parameters can change not

only over time but also over the states of the system and uses Extended Kalman filtering to estimate these parameters. The proposed method is called Bayesian SSA.

4. A spatial SSA model produces superior forecasts than a multivariate model based on all time series even in the presence of time series with modes that differ by a small cosine distance.
5. Projection pursuit using Joint Diagonalisation provides no improvement over the SVD when applied to Horizontal SSA.
6. We introduce a novel covariance clustering algorithm based on the self organising feature map which seeks to separate time series based on their SSA covariance structures. The method is called Self-Organising Eigenspace Map (SOEM). It is shown empirically that the consequent clustering produces superior forecasts and the mapping is topologically ordered.

1.3 Thesis structure

The thesis is organised as follows. We begin with a comprehensive literature review (Chapter 2) which is intended to service the subsequent chapters in the sense that all prior work is deposited there and is referenced when required. One exception is the inner workings of SSA which we give in Chapter 4 as these are required explicitly for a theorem.

Chapter 3 performs a preliminary analysis of SSA applying it to some well known data sets and in addition we examine bootstrapping when applied to the coefficients (previously having only been applied to the forecasts themselves). One aspect in particular, the performance of SSA with respect to a time series which has a structural break is found to be flawed and this leads into the next Chapter.

Chapter 4 develops a State Dependent Model (SDM) for application with the evolution of the states within SSA. Specifically, it is proven that standard SSA is flawed in the presence of a structural break. An SDM is then tested on synthetic data and real data in two scenarios, the univariate case and the multivariate case. During examination of the multivariate case it becomes clear that grouping time series together in a multivariate model is an important issue which leads into the next Chapter.

Chapter 5 examines latent and reduced space multivariate SSA time series modelling. The core question of this Chapter is how to determine which time series may or may not reinforce each other without performing an exhaustive search (which would require

construction and comparison of $N(2^{N-1} - 1)$ models). In this Chapter, we introduce a Spatial SSA model based on geographical information about the time series. In addition, we examine a means of clustering the data from the information contained in the time series themselves. In particular, we introduce a novel method similar to the self organising map which is based on matrix inputs rather than vectors.

Finally, in Chapter 6, conclusions are drawn from the previous chapters bringing together all the findings to demonstrate how SSA may be utilised especially in the presence of a structural break. This leads into an exploration of future directions for research.

1.4 List of publications and conference papers

The following publications leading from the work in this thesis have been submitted, published in full or are in preparation:

1. Fay, D. and Rahmani, D., (2017), "Clustering time series based on their singular spectrum covariances using a self organising eigenspace map", (in preparation).
2. Rahmani, D., Fay, D., Heravi, S. and Yoo, P., (2016), "Forecasting multivariate time series in the presence of structural breaks via Bayesian singular spectrum analysis using state dependent parameters", *European Journal of Operational Research* (submitted).
3. Rahmani, D. and Fay, D., Hui, P. (2016), "A tensor based singular spectrum analysis algorithm for multivariate time series with application to city wide vehicular traffic", (in preparation).
4. Heravi, S., Rahmani, D., Hassani, H. Fay, D. (2016), "Forecasting time series with structural breaks with Singular Spectrum Analysis, using a general form of recurrent formula", *Special issue on singular spectrum Analysis of international journal of forecasting*, (submitted).
5. Rahmani, D. and Callaway, A., (2016), "A comparison of models when forecasting limited data sets: A case using Olympic Archery", *journal of sport science*, (submitted).
6. Rahmani, D., "General Singular Spectrum Analysis Forecasting using State Dependent Models", (2016), the World Meeting of International Society for Bayesian Analysis in Italy.

7. Rahmani, D., "Multivariate Spatial Forecasting for the Contiguous United States", (2015), presented at 8th world congress in European Meeting of Statisticians in Amsterdam.
8. Rahmani, D., "Predicting Recessionary Times With Singular Spectrum Analysis Technique", (2015), presented at 42th AIB-UKI conference (Academy of International Business) in Manchester.
9. Rahmani, D., "A forecasting algorithm for Singular Spectrum Analysis based on bootstrap Linear Recurrent Formula coefficients", (2014), International journal of Energy and Statistics, 02, 287.
10. Rahmani, D., "A New Singular Spectrum Analysis Confidence Interval", (2014), presented at 34th International Symposium on Forecasting in Rotterdam.

Chapter 2

Background

2.1 Singular Spectrum Analysis

Singular Spectrum Analysis is a linear analysis and prediction method for a time series, in which its data-adaptive characteristics make it a suitable method for the spectral analysis of short and noisy time series, without any a priori knowledge of the process generating the data (Vautard et al., 1992). SSA is based on the principles of classical time series analysis, multivariate statistics, multivariate geometry, dynamic systems and signal processing. It was first introduced into chaos theory by (Broomhead & King, 1986) and (Fraedrich, 1986b) to explain the intrinsic dynamics of a time series. Afterwards it received significant scholarly attention in the literature from (Vautard et al., 1992; Allen & Smith, 1996) and (Golyandina et al., 2001). More investigations into both the theoretical and practical foundations of SSA can be found in (Zhang & Hui, 2012), (Zhao et al., 2011), (Kapl & Mueller, 2010), (Oropeza & Sachchi, 2011), (Yiou et al., 2000) and (Groth & Ghil, 2011), (Danilov & Zhigljavsky, 1997), (Golyandina et al., 2001), (Patterson et al., 2011) and the references therein. An introductory explanation of SSA may be found in (Elsner & Tsonis, 1996).

Over the last two decades, SSA has been recognised as a standard tool in a variety of fields such as climate and geophysics (Hou et al., 2014; Oropeza & Sachchi, 2011; Le Bail et al., 2014; Chen et al., 2013; Vautard & Ghil, 1989; Kondrashov & Ghil, 2006; Fraedrich, 1986a; Chang et al., 2015; Hou et al., 2014), engineering (Chao & Loh, 2014; Liu et al., 2014), medicine (Aydn et al., 2011; Thuraisingham, 2013; Ghaderi et al., 2011; Sanei et al., 2011) and many others; see (Muruganatham et al., 2013; Daly et al., 2013; Golyandina et al., 2001; Elsner & Tsonis, 1996; Qadrdan et al., 2013; Lisi et al., 1995). SSA has also been extended and modified in different aspects such as Toeplitz SSA (Golyandina et al., 2001), Monte Carlo SSA (Allen & Smith, 1996), sequential SSA (Golyandina et al., 2001), SSA based on minimum variance (Hassani, 2010), SSA based on perturbation (Hassani et al., 2011b) and multivariate SSA (Broomhead & King,

1986). In addition, (Zhigljavsky, 2010) studied SSA-based methods and their connection with sub-space models in signal processing.

In addition to forecasting, SSA is often used for smoothing time series; finding trends of diverse resolutions; finding structure in short time series; extraction of seasonal components and periodicity with varying amplitudes; simultaneous extraction of cycles with small and large periods; and simultaneous extraction of complex trends and periodicity (Golyandina et al., 2001). Furthermore, a study by Thomakos (Thomakos, 2008) proposes an asymptotically optimal linear filter for smoothing and trend extraction for unit root processes using SSA. Thomakos found that the SSA-based filter can provide similar performance to the Hodrick-Prescott filter (Hodrick & Prescott, 1997) and is thus an alternative method for extracting the cyclical component.

SSA also provides insight into the unknown or partially known dynamics of an underlying system (as detailed in Section 2.1.1 and subsequent sections) by using an embedded representation of the time series to decompose them into a set of data-adaptive orthonormal components. These components can be projected essentially into a lower dimension and then reconstructed to form a smoother time series, which can be used for explaining structure and used subsequently for forecasting. Neither normality nor stationarity are required for the time series which makes SSA a model-free technique with a broad range of applicability, (Golyandina et al., 2001).

A direct extension of SSA, multivariate or multichannel SSA (MSSA), has an ability to reconstruct shared dynamic and oscillatory behaviour of stochastic systems in terms of finding common oscillatory modes. MSSA, as a robust way of analysing the spatio-temporal behaviour of short and noisy time series, can greatly help with phase synchronisation analysis (Groth & Ghil, 2011). MSSA helps a time series analyst to extract common spectral components from the multivariate data set, along with the co-movements of different channels (Groth & Ghil, 2011). It is assumed that common properties exist which can explain the simultaneous variation in a system of multiple time series (Viljoen & Nel, 2010).

Essentially, MSSA can be applied for different purposes such as centralising feature extraction, extraction of shared behaviour, identification of coherent spatio-temporal structures (given a regular sampled archive of maps), explaining co-movement and capturing spatio-temporal dependence (Groth & Ghil, 2011; Golyandina & Zhigljavsky, 2013). For example, (de Menezes et al., 2014) combines SSA with periodic autoregressive models (PAR), to forecast the monthly average wind speed of two regions of Northeast Brazil (Hydrological time series). The outcome shows PAR(P)-MSSA is far superior than the two others (PAR(P) and PAR(P)-SSA), since MSSA takes into account the spatial dependence between the two stations.

Moreover, MSSA as a unified and robust method can be used for network anomaly detection. It is shown by (Babaie et al., 2014) that MSSA is able to detect a much

wider range of anomalies, regardless of their type, when compared to the Kalman filter and a Wavelet transform. MSSA could also capture temporal changes in traffic patterns as well as changes in the number of flows (Babaie et al., 2014).

MSSA is also useful for simultaneous analysis and forecasting of several time series. (Patterson et al., 2011) shows that MSSAs flexibility can be beneficial for forecasting real-time data subject to a revision process, when compared to those which involve restrictive assumptions such as linearity, normality and stationarity. In addition, applying bivariate SSA to UK industrial production indicators was found to improve the ability of the model over standard parametric models like VAR and ARMA. They demonstrate that MSSA has advantages over both state-space models and standard cointegration models in two aspects: first, its ability to automatically detect non-linear cross correlations, and second, its direct extension to high-dimensional systems both in theory and practice.

Additionally, MSSA can be used to reconstruct the entire attractor of a nonlinear dynamical system from limited data. For example, (Groth et al., 2011) shows MSSAs ability to separate distinct spectral components in a multivariate data set of limited length, in the presence of relatively high noise levels. The performance of MSSA and SSA in comparison with other classical models like autoregressive models (ARIMA and GARCH) and random walk models are examined through a wide range of financial and economic time series. For example, forecasting the US inflation rate using the consumer price index and gross domestic product (GDP) in (Hassani et al., 2013b), examining the volatility of floating currencies, and the UK and EU daily exchange rates in (Zhang & Hui, 2012), predicting market steel prices with real GDP for the USA, Germany and China in (Kapl & Mueller, 2010), forecasting Romanian exchange rates to the Euro in (Georgescu & Delureanu, 2015), modeling and forecasting the overall fluctuation of the USA unemployment rate in (Skare & Buterin, 2015), and in particular, forecasting the industrial production index by (Heravi et al., 2004; Patterson et al., 2011; Hassani et al., 2012; Osborn et al., 1999; Franses & Dick, 2000).

The basic implementation of MSSA (or SSA) for a time series analysis has two steps: Decomposition and Reconstruction. Each has two sub-steps: *Embedding and Singular Value Decomposition* and *Grouping and Diagonal Averaging*. The following section will present a description of the algorithm and related works (by considering each stage) in more general terms while more technical details will be explained in Chapter 4.

2.1.1 Decomposition

Decomposition is often a standard procedure in time series analysis. It provides insight into the mechanisms producing the time series patterns and behaviours. In fact, the key

idea behind the decomposition methods is that they capture the components which are physically meaningful (Bonizzi et al., 2014).

2.1.1.1 Embedding

The modeling of a deterministic dynamical system relies on the concept of embedding, as a collection of possible system states. An individual realisation of a dynamical system is the outcome of interacting variables, and thus ought to hold information about the dynamics of all the key variables interacting in the system. According to (Ruelle, 1980) instead of a continuous variable and its derivatives, a discrete time series with its successive shifts by a lag parameter ought to be enough to approximate the dynamics of a system. Consequently, lagged copies of the time series are considered as to explain the periodic content of the series under study. There exists a relationship between the lagged copies if Takens' embedding theorem is applicable. In principle, because of the embedding theorem (Takens, 1981), the time delayed version of a time series is generically sufficient to reconstruct the dynamics of the underlying systems if enough delayed coordinates are used (Cao et al., 1998).

SSA, as an embedding time series method, uses these lagged or delayed copies to analyse the time series. The procedure is also referred to as the method of delays, in which a single time series record, y_1, \dots, y_N , is embedded into a multivariate set of delayed records, $\mathbf{X}_i = [y_i \dots y_{i+L}]^T$, $i = 1, \dots, K$, where $K = N - L + 1$. The trajectory matrix of the series is then given by $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_K] \in \mathcal{R}^{K \times L}$. The trajectory matrix includes the complete record of repeating patterns that have appeared within a window size L . The window length is also known as the *embedding dimension*.¹ In principle, the trajectory matrix can be examined for repeating patterns that are representing trends and oscillations in the original time series. Specifically, these patterns are examined via the covariance matrix of the trajectory matrix, i.e. $\mathbf{X}\mathbf{X}^T$ matrix. The elements of $C = \mathbf{X}\mathbf{X}^T$ matrix are proportional to the linear correlation between all pairs of series used to construct \mathbf{X} .

There are several different ways to construct C from a time series (Elsner & Tsonis, 1996). Among the most frequently used are via construction of Hankel or Toeplitz matrices. The Hankel based approach uses the data directly such that given a set of observations, y_1, \dots, y_n , and a fixed window length L , the elements of the Hankel matrix, using the product of the trajectory matrix and its transpose, are $c_{ij} = \frac{1}{N-L+1} \sum_{t=1}^{N-L+1} y_{i+t-1} y_{j+t-1}$ where $i, j \leq L$. This matrix is originally used by (Broomhead & King, 1986) which consists of a basic SSA. Alternatively, (Vautard & Ghil, 1989) used a Toeplitz structure to compute the *lagged-covariance matrix* elements. There are different ways to estimate lagged-covariance matrices elements such as the Yule-Walker

¹ The embedding dimension ensures that all synchronisation between the lagged channels are taken into account (Pukenas, 2014).

or Burgs algorithm, principal component analysis (PCA), etc. see (Vautard et al., 1992). A robust estimate of the lagged-covariance matrix considered in many studies is given by: $c_{ij} = \frac{1}{N-|i-j|+1} \sum_{t=1}^{N-|i-j|+1} y_{|i-j|+t} y_t$. In (Vautard et al., 1992) they show that there is a bias-variance trade off when there are very low frequency observations in the system, or when the number of observations is small. They also presented a particular method to estimate the Toeplitz matrix which has little bias compared to others. ²

Note that the Toeplitz matrix as defined above resembles an auto-covariance matrix where all elements along each of the diagonals are equal. Therefore, drawing inferences from the spectral decomposition of a covariance matrix with the Toeplitz structure implies implicitly an assumption of stationarity for the time series. It is shown by (Elsner & Tsonis, 1996) that the eigenvectors of the Toeplitz matrix captures less of the variance in a times series as compared with a Hankel matrix. They show that the eigenvectors of the Hankel matrix span the entire variance spectrum more precisely than a Toeplitz matrix. It is therefore obvious that the Toeplitz lagged covariance matrix cannot successfully explain the variance of a trend which is sometimes an important source of error in prediction algorithms. On the other hand, unlike the Toeplitz form, the Hankel matrix is more robust, regardless of the underlying physical process from which the time record was sampled (Elsner & Tsonis, 1996). More details on the application of both Toeplitz and Hankel matrices can be found in (Golyandina et al., 2001).

The above procedure can be extended for a multivariate time series. Specifically, for M different time series, there exist M different $L_i \times K_i$ trajectory matrices, $\mathbf{X}^{(i)}$ ($i = 1, \dots, M$), and these can form a *stacked* Hankel matrix in either a horizontal or vertical format (Hassani & Mahmoudvand, 2013). The former version enables us to have various K_i and different series length N_i , however, there is an equal window length L_i for all of the series. The result of this step is as follows:

$$\mathbf{X}_H = \begin{bmatrix} \mathbf{X}^{(1)} & \mathbf{X}^{(2)} & \dots & \mathbf{X}^{(M)} \end{bmatrix},$$

where, \mathbf{X}_H indicates that the output of the first step is a stacked Hankel matrix formed in a *horizontal* form. It is directly apparent from the structure of the matrix $\mathbf{X}_H \mathbf{X}_H^T$ in a horizontal format, that there is not any cross-product between Hankel matrices $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ and rather that the sum of $\mathbf{X}^{(i)} \mathbf{X}^{(i)T}$ ($i = 1, \dots, M$) provides the stacked Hankel matrix.

To form the stacked Hankel matrix in a vertical form, it needs to have $K_1 = \dots = K_M = K$. Accordingly, this version enables us to have various window lengths L_i and different series lengths N_i , however K_i is restricted to being equal for all of the series. The result of this step is the following stacked Hankel matrix:

²These two forms of matrices can be transferred to each other by using a *backward identity* permutation P where PT is a Hankel matrix for any Toeplitz matrix T and PH is a Toeplitz matrix for any Hankel matrix H .

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

The structure of the matrix $\mathbf{X}_V \mathbf{X}_V^T$ is identical to the variance-covariance matrix in classical multivariate statistical analysis. For the i^{th} series, the matrix $\mathbf{X}^{(i)} \mathbf{X}^{(i)T}$ appears along the (block) diagonal and the products of two Hankel matrices $\mathbf{X}^{(i)} \mathbf{X}^{(j)T}$ ($i \neq j$), which are related to the series i and j , appears in the off-diagonals.³

2.1.1.2 Projection pursuit

The basic idea behind a multivariate analysis is that some of the information about the data is redundant, and the key characteristic of the data can then be explained in terms of their tendency to concentrate into clusters, or about a curve or non-flat surface (Friedman & Stuetzle, 1982). The term *projection pursuit* was first introduced by (Friedman & Tukey, 1974) as a technique to map multivariate data onto a lower dimensional manifold. Projection pursuit reveals patterns of variation in the given data set by proposing a low-dimensional orthogonal projection of it to find *interesting directions*. While projecting from the lower dimensions the appearance of the projected data set does not change abruptly as the projection direction varies, and the space of projection directions is of a low dimensionality (Friedman, 1987). On the other hand, for projection from higher dimensions, it can still be true that the appearance of the projected data set changes smoothly, however, it becomes impractical to explore potential comprehensive projections due to the high dimensionality of the space of projection directions (Tukey & Tukey, 1981). (Friedman & Stuetzle, 1982) discussed the interactive process for such an exploration, although (Friedman, 1987) showed that those processes are unable to improve the applicability of the approach as far as it is needed. Thus, an automatic procedure for selecting potentially interesting projections was essentially needed. Projection pursuit can provide such selections by locally optimising the projection directions according to some measure of what (Friedman, 1987) calls *interestingness*.

A projection pursuit algorithm associates with each and every direction in the multi-dimensional space, a continuous index which measures its interestingness as a projection axis, and then varies the projection direction so as to maximize this index over the projection space. Basically, the projection index measures how much structure is contained within orthogonal linear projections of the data (Jones & Sibson, 1987). The choice of index is the most crucial part in projection pursuit. In (Friedman & Tukey, 1974) the index is formed as a product of a robust measurement of scale (like variance) with

³Note that if X with dimension $LM \times K$ denotes the vertical trajectory matrix, then X^T is the horizontal trajectory matrix with dimension $K \times ML$ (suppose all series have similar length and similar window length).

a weighted measure of a number of close pairs (Huber, 1985). This locally optimal projection, it is proposed, can give interesting insights into the data.

Suppose Y is an $M \times N$ matrix of data. The projected data, Z , is then formed by $Z = A^T Y$ where A is a linear map. The projection index, I , is usually expressed as: $I(Z) = I(A^T Y) = I(A)$. In the distributional format of a projection pursuit, it is assumed that the projected data, Z , have a density function, f , that essentially depends on A . Therefore, the most practical procedure to compute a projection index is by considering the density function of Z , and the projection index can be then written as $I(f)$. Another format which is called the *sample case* can be carried over by replacing the actual density function with an empirical estimate (Friedman & Tukey, 1974).

To optimise the projection index, the hill-climbing optimisation method is used by (Friedman & Tukey, 1974) to find interesting projections, which are $I(A) = S(A)D(A)$, whereas $S(A)$ measures the general spread of the data, and $D(A)$ measures the local density of the data after projection onto a projection vector. There are numerous specific ideas for revealing interesting projections, such as the use of entropy as the basis for an index function (Jones & Sibson, 1987), centring and sphereing (Tukey & Tukey, 1981), projection indices based on order-entropy (Rényi, 1961), moment indices and the Friedman index (Friedman, 1987), Halls index (Hall, 1989), the Morton index (Morton, 1989) etc. (Jones & Sibson, 1987) concentrates specifically on the practical implementation and application of the projection pursuit, whereas (Huber, 1985) explores the projection pursuit framework in more detail and relates to the projection pursuit derivatives as mentioned above.

Essentially, projection pursuits methods are based on an interpretation of the eigenvectors of the covariance matrix as directions that maximize or minimize variances. By replacing the variance by other indices, a variety of methods are developed. As will be seen, the Singular Value Decomposition (SVD), principal component Analysis (PCA), Independent Component Analysis (ICA), Joint Diagonalisation (JD), and the Tucker Decomposition are several popular particular implementations of a projection pursuit algorithm.

Singular Value Decomposition The lagged-covariance matrix of MSSA, C , is real and symmetric, and then there is a diagonalizing matrix with orthonormal columns, U , in which $C = U^T \Sigma U$.⁴ As a spectral decomposition of C , it consists of a summation of the dimensional projections of eigenvectors, $U_i U_i^T$. The diagonal matrix Σ consists of ordered eigenvalues, and their square roots are called the singular values, λ , of \mathbf{X} .

$$C = U^T \Sigma U \Rightarrow \mathbf{X} \mathbf{X}^T U^T = U^T \Sigma \Rightarrow U \mathbf{X} \mathbf{X}^T U^T = \Sigma \Rightarrow (U \mathbf{X})(U \mathbf{X})^T = \Sigma \quad (2.1)$$

⁴I.e. the left and right singular vectors are equal

$U\mathbf{X}$ is the trajectory matrix projected onto the orthogonal basis U . The components of \mathbf{X} ordered along with the basis U are uncorrelated, since U constitutes orthogonal vectors called singular vectors of \mathbf{X} .

Based on (Elsner & Tsonis, 1996), the name singular spectrum relates to the spectral (eigenvalues) decomposition of the matrix into a set (spectrum) of eigenvalues. These eigenvalues are the numbers that result in a matrix $X - \lambda I$, which is singular. The traditional eigenvalue decomposition approach, which involves multivariate data to analyse the singular spectrum, might be mistaken for SSA. Hence, more precisely, singular spectrum analysis (SSA) is the analysis of the time series using the singular spectrum.

The link between SSA and other classical spectral analysis techniques is investigated by (Vautard & Ghil, 1989). Their study shows that when a sharp peak occurs in the power spectrum, a data adaptive choice of filters gives a fundamentally more adaptable tool than from a standard spectral analysis. It means the data-adaptive characteristics of an SSA basis (its eigen-components) gives SSA a significant advantage over classical spectral analysis methods (where the basis function is prescribed as sines and cosines).

Principal component analysis SSA is closely related to the classical and commonly used principal component analysis (PCA) (Elsner & Tsonis, 1996). The former provides truly dynamical information, while the latter gives mostly a geometrical description (Vautard & Ghil, 1989). The key idea in both cases is to identify the main patterns of covariance matrix variability, in decreasing order of the associated variance. (Plaut & Vautard, 1994) showed that apart from the difference in defining the trajectory matrix, there is no difference between the expansion used in classical PCA and the expansion used in SSA.

Basically, PCA uses a covariance matrix of normalised data (centred and scaled to have unit variance) to extract the principal eigenvectors (Jolliffe, 2002; Abdi & Williams, 2010). Those eigenvectors are stacked as P , a rotation matrix in descending order. Its multiplication by C provides the data for PCA. New variables generated by PCA included useful information from the original data, while redundancy has been removed. PCA basically supposes that the direction with the largest variance corresponds with the principal dynamics of a system and the remainder corresponds to redundant information. It can also be chosen by setting a threshold for the variance, or by choosing the number of significant square root of eigenvalues (Jolliffe, 2002).

There is a simple relationship between PCA and the SVD. PCA is equivalent to performing the SVD on the centred data, furthermore, the right eigenvector, matrix V , of the SVD is equivalent to the rotation matrix returned by PCA (Ghil et al., 2002). Note that, the diagonal elements of Σ from the SVD are proportional to the standard deviations returned by PCA. The elements of Σ are formed by taking the sum of the

squares of the principal components, but not dividing them by the sample size. Studies like (Vautard et al., 1992) used PCA instead of SVD in SSA.

It is worth noting that, the inclusion of temporal correlations makes MSSA superior to PCA in the extraction of dynamical behaviour (Groth & Ghil, 2011). A modified version of MSSA is introduced by (Groth & Ghil, 2011) by applying variance-maximization rotation to MSSA eigenvectors to optimally detect clusters of a synchronised oscillator. Moreover, heteroscedasticity between symmetric covariance matrices is also taken into account by (Viljoen & Nel, 2010) through the common principal component (CPC) approach. CPC decomposes the series into the sum of a common small number of components, which are related to a common trend and to oscillatory components and noise (common base vectors). It focuses on the uncorrelatedness of PCs rather than on the aspect of maximising the amount of variability accounted for in the predefined number of principal components.

Independent component analysis When several components of the original time series are mixed in such a way, that their contributions are very similar, then the optimality of the SVD does not help to separate these components properly. This is due to the lack of strong separability between components. (This will be explained in section 2.1.3.) In this situation, special rotations ought to be found in such a way that they satisfy some additional optimality criterion (Golyandina & Zhigljavsky, 2013). ICA-SSA allows SSA to deal with deterministic components, while ICA is a form of projection pursuit that seeks a non-Gaussian distribution in the projected data (Golyandina et al., 2001).

Unlike PCA, which minimises the variance directions in the covariance matrix of the series, ICA minimises higher order statistics like kurtosis-based measures. The covariance matrix, C , is assumed to be a linear combination of non-Gaussian and statistically independent components, in which $C = SA$ where columns of S contain the independent components, and A is a linear mixing matrix (Hyvärinen & Oja, 2000). ICA then looks for an un-mixing matrix W to minimise mutual information, which is equivalent to maximising the non-Gaussianity under the condition that W is an orthonormal matrix. To measure non-gaussianity FastICA relies on neg-entropy $J(y) = [E\{G(y)\} - E\{G(\nu)\}]^2$; $\nu \sim N(0, 1)$, which is more robust than kurtosis-based measures and is fast to compute (Hyvärinen & Oja, 2000). (Hyvärinen & Oja, 2000) suggest the following choices for G as a non-quadratic function: $G(u) = \frac{1}{\alpha} \log \cosh(\alpha u)$, and $G(u) = \exp(u^2/2)$ where $1 \leq \alpha \leq 2$.

It is worth pointing out that for a time series with a shared common mode (cointegrated series), which explains the variation between different time series, ICA as a pre-processing step before SSA can lead to a better decomposition than the SVD. To prove that, (Chan, 2013) demonstrates that projections that are maximally distant from

the Gaussian distribution can be useful in identifying a cointegration test based on ICA. However, the fundamental limitation in ICA is that the independent components must be non-Gaussian for ICA to be possible. Note that ICA is a less stable procedure than MSSA therefore it is suggested that it is not a good idea to replace the SVD completely with ICA (See (Golyandina et al., 2001) on the use of ICA with MSSA).

On the other hand, applying SVD to a large covariance matrix can make MSSA computationally expensive. This issue is addressed by (Pukenas, 2014) via reordering the Toeplitz block covariance matrix into a block Toeplitz matrix. Then this matrix is embedded into a block circulant matrix, and is efficiently block-diagonalized by means of the Fast Fourier Transform. This combination results in a version of MSSA which is less computationally complex for the detection of smooth changes in large spatially extended systems. However, when multiple stochastic processes mixed with periodic components occurs simultaneously, sparse dynamical characteristic of non-stationary time series like structural breaks and shifts cannot be tracked directly.

Joint Diagonalisation Specifically, Joint Diagonalization (JD) defines an *average eigenstructure* shared among multiple covariance matrices. JD extends the eigenvalue-eigenvector decomposition (diagonalization of a single matrix) and its generalised version (diagonalization of a matrix pair) to a matrix set including three or more matrices. Therefore, the resulting decomposition is considered as an extension of the SVD to a stack of matrices. Joint Diagonalization was introduced by (Bunse-Gerstner et al., 1993) for simultaneous diagonalization of multiple matrices by mainly contributing to the stability and convergence concern. The simultaneous diagonalization algorithm is based on an extension of the Jacobi matrix: a joint diagonally transformation which is iteratively optimised under plane rotations (Cardoso & Souloumiac, 1996). The joint diagonalization technique can be used as a generic algorithmic tool for various applications, such as blind source separation (Molgedey & Schuster, 1994; Belouchrani et al., 1997; Pham & Cardoso, 2001; Yeredor, 2002; Ziehe et al., 2004), common spatial pattern analysis (Koles, 1991; Blankertz et al., 2008), common principal component analysis (Flury, 1984; Fengler et al., 2003), signal processing (van der Veen et al., 1992; Van der Veen et al., 1998) and, more recently, kernel-based nonlinear blind source separation (Harmeling et al., 2003). It is shown that for more than two commuting matrices it may not be possible to achieve joint diagonalization with one single transformation. However, exact joint diagonalization of more than two matrices is applicable if the matrices have a certain common mode, as it is the case for the multivariate time series application. If there is not such a common structure, then one can only use an approximate joint diagonalization. The goal of a joint diagonalization algorithm is to find a matrix U that simultaneously set off diagonal terms of C_1, \dots, C_M to be zero by a unitary transform. This is the form of the joint diagonalization process that has been used most frequently

in the literature, for example in (Bunse-Gerstner et al., 1993; Cardoso & Souloumiac, 1993, 1996; Hori, 1999; Joho & Rahbar, 2002; Joho & Mathis, 2002; Fay & Yoneki, 2011).

The JADE algorithm is an implementation of JD for blind source separation which starts by estimating a whitening matrix, \hat{W} and set $Z = \hat{W}C$. Next a maximal set of cumulant matrices (statistics of an order higher than two), \hat{Q}_i^Z can be estimated to optimise an orthogonal contrast. Intuitively this means we seek a rotation matrix \hat{U} such that the cumulant matrices are as diagonal as possible, that is, solve:

$$\mathbf{J}_2(U) = \operatorname{argmin}_U \sum_i \operatorname{off}(U^T \hat{Q}_i^Z U) \quad (2.2)$$

They finally estimate an *unmixing* matrix, A , as $\hat{A} = \mathbf{J}_2(U)\hat{W}^{-1}$ and estimate the components of the source signal as $\hat{S} = \hat{A}^{-1}C = \mathbf{J}_2(U)^T Z$ as the source signal. The link with a joint-diagonalization criterion is the key for the derivation of the practical JADE algorithm (Moreau, 2001).

JD essentially seeks the rotation matrix, $U = \hat{A}^{-1}$, by minimising the diagonality criterion, $\mathbf{J}_2(U)$ where the $\operatorname{off}()$ is the Frobenius norm of the off-diagonal elements. The well-known Jacobi algorithm is used to find the joint diagonalizer. Moreover, the plane rotations are applied not only to the data but also to the cumulant matrices themselves, which means the JADE algorithm updates not only the data but the matrix-valued statistics as well. A key problem in these sort of algorithms is that the selection of the cumulant matrices should be involved in the estimation⁵. As shown by (Cardoso, 1999), the joint diagonalization criterion, $\sum_i \operatorname{off}(U^T \hat{Q}_i^Z U)$ is identical to the contrast function by using a maximal set of cumulant matrices. It can practically avoid the shortcomings of the previous joint diagonalization algorithms. However, (Cardoso, 1999) shows there is no other way for a priori selecting cumulant matrices in such a way that it still guarantees equivariant estimates, because the algorithm, although operating on statistics of the sphered data, also optimizes implicitly a function of $U^T Z$ only.

On the other hand, the algebraic nature of cumulant matrices is tensor based (McCullagh, 1987). It was first proposed by (Comon, 1997) as a Jacobi approach for ICA. (See (Comon, 1997) for a data-based algorithm and an earlier reference therein for the Jacobi update of higher order cumulant tensors.) Such a data-based Jacobi algorithm works through a sequence of Jacobi sweeps on the multi-way matrices until a given orthogonal contrast function is optimised.

Tucker Decomposition A Multidimensional time series can also be represented by a tensor. Simply speaking a tensor is just a multidimensional array of series which

⁵Basically the JADE algorithm is derived using fourth-order cumulants, therefore, its underlying contrast function is given with fourth-order cumulants as opposed to ICA, which remains available whatever the order of cumulants (Moreau, 2001)

preserves their true multiway structure (Kolda & Bader, 2008). Tensors are basically suitable for large scale data in pattern recognition, text mining, signal processing, computer vision, traffic analysis etc.

(Sun et al., 2008) describes a tensor of order M as a data cube with M dimensions. The most popular tensor decomposition algorithms can be considered as Tucker decomposition and PARAllel FACtor Analysis (or CANonical DEComposition) (PARAFAC/CANDECOMP). The former is a higher-order form of principal component analysis, while the latter is a higher-order extension of a singular value decomposition, which decomposes a tensor as a sum of rank-one tensors, (Kolda & Bader, 2008; Sun et al., 2008). Specifically, these higher order decomposition techniques capture multi-linear and multi-aspect structures in large-scale higher-order data-sets, which are key tools for feature extraction and supervised learning problems (Phan & Cichocki, 2010).

For the Tucker and PARAFAC/CANDECOMP decompositions there exist many different algorithms such as PARAFAC2, CANDELINC, DEDICOM, and PARATUCK2, as well as nonnegative variants of all of the above (Kolda & Bader, 2008). Most of them are essentially based on Alternating Least Squares (ALS) and Hierarchical ALS. Although, discussion about concrete implementations of those algorithms is out of the scope of this research. In what follows, this study will recall definitions of Tucker and PARAFAC/CANDECOMP. For more details see (Sun et al., 2008).

Definition 2.1 (Tucker Decomposition). . Given an M^{th} -order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_M}$ and core tensor sizes $\{r_1, \dots, r_M\}$, where $n_i (1 \leq i \leq M)$ is the dimensionality of the i^{th} mode, find a core tensor $\mathcal{Y} \in \mathbb{R}^{r_1 \times r_2 \times \dots \times r_M}$ and a sequence of projection matrices $\{U^{(d)}\}_{d=1}^M \in \mathbb{R}^{n_d \times r_d}$, such that $\|\mathcal{X} - \mathcal{Y} \times_1 U^{(1)} \times \dots \times_M U^{(M)}\|$ is small, that is, $\mathcal{X} \approx \mathcal{Y} \times_1 U^{(1)} \dots \times_M U^{(M)}$.

Note that the key idea behind the Tucker decomposition (Tucker, 1966) is that a big tensor can be approximated using a small tensor through a change of basis. The approximation can be exact if $r_d = n_d$ and $U^{(d)}$ is full rank for $1 \leq d \leq M$ (Sun et al., 2008). In fact, the basis varies for each mode through the projection matrix, $\{U^{(d)}\}_{d=1}^M$.

It would clearly have been sufficient to assume that the core tensor is superdiagonal which is the case for PARAFAC/CANDECOMP. This is a special case of the Tucker decomposition. In addition, PARAFAC/CANDECOMP as a generalisation of SVD for higher order arrays, is a weighted sum of rank one tensors.

Definition 2.2 (PARAFAC/CANDECOMP). Given a M -order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_M}$, we wish to find r rank-one tensors in the form of $\lambda_i \mathbf{u}_i^{(1)} \circ \dots \circ \mathbf{u}_i^{(M)}$ where $\{\mathbf{u}_i^{(d)}\}_{d=1}^M \in \mathbb{R}^{n_d}$ for $1 \leq i \leq r$ such that $\|\mathcal{X} - \sum_{i=1}^r \lambda_i \mathbf{u}_i^{(1)} \circ \dots \circ \mathbf{u}_i^{(M)}\|$ is small, that is, $\mathcal{X} \approx \sum_{i=1}^r \lambda_i \mathbf{u}_i^{(1)} \circ \dots \circ \mathbf{u}_i^{(M)}$.

Subjecting the factor matrices to the condition of orthogonality is yet another special case, such as Higher Order Singular Value Decomposition algorithm or the Higher Order Orthogonal Iterations algorithm, introduced by (Lathauwer, 2011). In the higher order SVD algorithm all modes are assumed to be independent and thus matrix SVD can be performed on each matricization of the tensor. On the other hand, the higher order orthogonal iterations algorithm performs an iterative process to seek out better projection matrices. In particular, the higher order SVD with one iteration is a special case of the higher order orthogonal iterations algorithm.

Recently, (Kouchaki & Sanei, 2013; Kouchaki et al., 2015) employed a PARAFAC/CANDECOMP instead of the SVD in SSA and termed this Tensor SSA. (A 3D-tensor is decomposed by PARAFAC/CANDECOMP.) It is shown that tensor SSA enables SSA to perform better for single channel data decomposition in nonstationary and underdetermined source separation, while more classical techniques like ICA cannot even be used directly (Kouchaki & Sanei, 2013). In (Kouchaki et al., 2015) tensor SSA (TSSA) is used for sleep EEG analysis and the results are compared with the clinical results, which confirms the ability of the method to achieve a better understanding of sleep EEG data. In (Kouchaki et al., 2015) empirical mode decomposition is used to select the subgroup of the desired signal as an adaptive supervised approach. The method is applied to both simulated and real data. Some samples of the narrow-band and non-stationary signal with low SNR were generated to test the performance of TSSA, which achieved the lowest RMSE in comparison to SSA and ICA. The real application is devoted to sleep EEG, which shows that the proposed method can capture the transitions between stages of sleep by more accurately evaluating brain activity variations.

Apart from the above-mentioned methods, there are other studies on expanding the decomposition part of SSA for application to various time series. For example, (Vautard et al., 1992) combine SSA with more advanced spectral-analysis methods - such as the maximum entropy method (MEM) and the multi-taper method (MTM) - to refine the interpretation of oscillatory behaviour (Vautard et al., 1992). Later on, (Yiou et al., 2000) extended Toeplitz SSA to a nonstationary time series by using multi-scale ideas from wavelet analysis. He suggests a moving window proportional to series length, and uses a wavelet transform to detect the regular part of the lag-covariance matrix. The method is called Multi-scale SSA (or MS-MSSA), which considers eigenvectors as data-adaptive wavelets to address the non-stationarity of the series. Alternatively, (Moskvina & Zhigljavsky, 2003) replaced the SVD with a spectral projector of the lag-trajectory matrix, which is the orthogonal projector onto the eigenspace for cutting off eigenvalues at a certain size. Their numerical examples show a high degree of polynomial approximation of the spectral projector that influences the quality of the reconstruction.

2.1.2 Reconstruction

The next step after decomposing the series is reconstruction. Clearly, those specific components of the series identified by the projection pursuits methods can now be used for the reconstruction of the time series. Reconstruction stage consists of two sub-steps: Grouping and Diagonal averaging.

2.1.2.1 Grouping

The grouping procedure partitions the set of indices $1, \dots, L$ into r disjoint subsets I_1, \dots, I_r . The process of choosing the sets I_1, \dots, I_r is called *eigentriple grouping*. The aim of this grouping step is the separation of the additive components of the time series.

Despite the fact that several formal criteria for separability can be introduced (as will be discussed in Section 2.1.3), the whole procedure of splitting the terms into groups (i.e., the grouping step) is difficult to formalise completely (Golyandina & Zhigljavsky, 2013). The principals and approaches of identifying the specific components for their inclusion into different groups are discussed mainly in (Pukenas, 2014; Golyandina & Zhigljavsky, 2013).

2.1.2.2 Diagonal Averaging

The final step in MSSA is to reconstruct the components of the original series. There is a formal procedure of transforming an arbitrary matrix into a Hankel matrix, and therefore, into a series which is called *diagonal averaging or hankelization*. It defines a matrix \mathbf{Y} with the values of the time series y_i as averages for the corresponding anti-diagonals of the matrices \mathbf{X}_{I_i} . Basically each \mathbf{X}_{I_i} can be seen as the Hankel matrix for the corresponding embedded component series. (More details about Hankelisations will be given in Chapter 4.)

Hankelization is an optimal procedure in the sense that the matrix $\mathcal{H}\mathbf{Y}$ is closest to \mathbf{Y} (with respect to the Frobenius matrix norm) among all Hankel matrices of the corresponding size (Golyandina & Zhigljavsky, 2013). In its turn, the Hankel matrix $\mathcal{H}\mathbf{Y}$ defines the series uniquely by relating the values in the anti-diagonals to the values in the series.

As stated before, there are two parameters in MSSA: the first parameter is the embedding dimension or window length L , and the second parameter is the number of components for reconstruction r . Choosing improper parameters yields an incomplete reconstruction and produces misleading results in forecasting. It is therefore pertinent to briefly comment on the differences between the historical approach and the relatively new, automated approach used to select those parameters.

2.1.3 Parameter Selection

Given that MSSA parameters are the crucial determinants underlying the performance of the MSSA, no theoretical solution has been yet proposed to solve this problem. Of course, there are worthwhile efforts and various techniques for selecting the appropriate value of L (see, for example, (Golyandina et al., 2001), (Golyandina, 2010), (Hassani et al., 2011a) (Sauer et al., 1991) and (Mahmoudvand et al., 2013)).

2.1.3.1 Embedding dimension

As mentioned before the trajectory matrix in the decomposition stage of MSSA is generated by windowing the time series and storing all windows in the matrix. The length of windowing, or the embedding dimension, must be properly selected, as the reconstructed components are strongly dependent on it. Even though guidelines have been provided on the choice of the window length, (Vautard et al., 1992; Golyandina et al., 2001; Golyandina & Usevich, 2004) confirm that there is no universal rule for the selection of the window length.

Certainly, the choice of parameter L is very much dependent on the given data, and also, for the analysis which it aims to perform. An improper choice of L would imply an inferior decomposition. Some discussions are given by (Elsner & Tsonis, 1996), and this author remarked that selecting L as being equal to a quarter of the length of a given series is a common practice. However, previously in (Golyandina et al., 2001) it was noted that L should not exceed half of a given time series. Large values of L allow longer period oscillations to be resolved, however, choosing L too large leaves too few observations from which to estimate the covariance matrix of the L variables.

It must also be noted that variations in L may influence both weak and strong separability features of SSA, i.e., both the orthogonality of the appropriate subseries of the original series and the closeness of the singular values (Golyandina & Zhigljavsky, 2013). For example, the weighted correlation between the signal and noise component has been proposed in (Golyandina et al., 2001) to determine the suitable value of L in terms of separability.⁶

In (Hassani, 2007) an analysis of the periodogram is determined to find out any strong signals (i.e. seasonal fluctuations) in the data set. Thereafter, one selects L proportionate to the seasonal fluctuations after which an analysis of the scree plot or paired eigenvectors enables this study to differentiate between signal and noise. Other approaches for the selection of SSA parameters are presented in (Hassani et al., 2011a,b; Khan & Poskitt, 2013) where the authors consider the selection of L based on the concept of separability between signal and noise. In addition, when forecasting during a

⁶A detailed discussion on this topic can be found in (Golyandina et al., 2001) section 1.6 and (Golyandina, 2010))

recession, or immediately following the impact of a major structural break, in (Hasani et al., 2013a) it was shown that a small trajectory matrix approach, whereby L is considered to be equal to 2, can provide better forecasts.

There are some versions of Basic SSA where the window length L is chosen automatically (see, for example, (Mahmoudvand et al., 2013)). A theoretic analysis of the signal-noise separation problem in SSA has also been considered in (Khan & Poskitt, 2013). Although considerable attempts and various techniques have been considered for selecting the proper value of L , there is not enough algebraic and theoretical material for choosing L . There are several criteria for determining window length that can be categorised into two groups as follows:

1. Criteria that consider different features of SSA without taking into account the type of data (general-based criteria), and
2. Criteria that consider the aim of the analysis and depend on the type of data (problem-based criteria).

According to these criteria, some recommendations are given for choosing window length (Vautard et al., 1992; Golyandina et al., 2001; Golyandina & Usevich, 2004). It is clear that large values of L gives more components and may increase the chances of separability. On the other hand, K can be considered as the number of L -variate samples from an original time series, and thus large values of K can be considered as the better choice from a statistical point of view. These two aims can be obtained when $L \times K$ is maximized, which is the number of observations in the trajectory matrix. Now, it is easy to see that $L = L_{\max}$ gives maximum value of the number of entities of the trajectory matrix.

Recently, the Singular Spectrum Decomposition (SSD) is introduced by (Bonizzi et al., 2014) which automatically select the window length (the embedding dimension) and the principal components of the trajectory matrix. In the SSD method, the choice of the embedding dimension, and the selection of the principal components for the reconstruction of a specific component series have been made fully data-driven. Moreover, a new format of the trajectory matrix is proposed, which enhances the oscillatory content in the data and guarantees the decrease of energy of the residual (Bonizzi et al., 2014).

Separability (w-correlation) The main concept in studying SSA properties is separability. As mentioned earlier, very helpful information for detection of separability and group identification is contained in the so-called w-correlation matrix. Well separated components have small correlation, whereas, poorly separated components generally have large correlation (Golyandina & Shlemov, 2013). Therefore, looking at the matrix of w-correlations between elementary reconstructed series one can find groups

of correlated series components and use this information for the consequent grouping (Golyandina, 2010).

The following criteria (called *w-correlation*) is a standard measure of similarity between two series $Y_N^{(1)}$ and $Y_N^{(2)}$ (Golyandina et al., 2001):

$$\rho_{12}^{(w)} = \frac{\langle Y_N^{(1)}, Y_N^{(2)} \rangle_w}{\|Y_N^{(1)}\|_w \|Y_N^{(2)}\|_w}, \quad (2.3)$$

where $\langle Y_N^{(i)}, Y_N^{(j)} \rangle_w = \sum_{n=1}^N w_n^{L,N} y_n^{(i)} y_n^{(j)}$, $(i, j = 1, 2)$, $\|Y_N^{(i)}\|_w = \sqrt{\langle Y_N^{(i)}, Y_N^{(i)} \rangle_w}$ and

$$w_i = \begin{cases} i + 1, & 0 \leq i \leq L^* - 1, \\ L^*, & L^* \leq i < K^*, \\ N - i, & K^* \leq i \leq N-1. \end{cases}$$

where $K^* = \max(L, K)$ and $L^* = \min(L, K)$.

If the absolute value of the *w*-correlations is small, then the corresponding series are almost *w*-orthogonal, however if it is large, then the two series are far from being *w*-orthogonal and are therefore, weakly separable. (Hassani et al., 2011b) showed the minimum value of *w*-correlation can be achieved at $L = L_{\max}$, for a wide class of time series. In SSA terminology, Y_1 and Y_2 show the components that are provided after the grouping steps. As mentioned, if two reconstructed components have zero *w*-correlation, it means that these two components are separable. Large values of *w*-correlations between the reconstructed components indicate that the components should possibly be gathered into one group and correspond to the same component in the SSA decomposition.

As discussed before, the window length L is the only parameter in the decomposition stage. Theoretical and empirical results confirm that L should be large enough, however, it should not be greater than $N/2$. Thus, based on this window length and on the SVD of the trajectory matrix, there should be a maximum number of $N/2$ eigentriples, ordered by their contributions (shares) into the decomposition stage. 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, a means of identifying them is required as now discussed.

2.1.3.2 Selection of components for reconstruction

A proper SSA decomposition can capture three main components: trends, oscillatory components (such as seasonality), and noise. Yet another challenge in MSSA is the

identification of the leading SVD components of the trajectory matrix. The key condition for separability is then that the deterministic part of the series can be approximated by a time series of finite rank r (Golyandina, 2010). This means that all except the r leading eigenvalues are close to zero.

At this stage one would select the appropriate number of eigenvalues r for reconstruction and consider the remainder as noise. Whilst this task would be simple in the case of a small time series, it becomes increasingly complicated and difficult when one has to analyse a huge number of paired eigenvectors for a larger time series. Moreover, in the absence of seasonal fluctuations the selection of L and r would be even more difficult, and in such cases the starting point is to select L such that it is less than half of the series length. According to (Golyandina & Zhigljavsky, 2013) the proper grouping needed, to obtain a suitable series decomposition, can be impossible if the signal components are mixed. For example, if the eigenvector contains periodic components with a slowly varying trend, this means that the trend and periodic components are not separable, at least for the chosen window length L . If it appears that for the chosen L there is no separability (weak or strong), the attempt to obtain separability is performed with other choices of L (Golyandina & Shlemov, 2013). For example, a possible lack of strong separability between a trend of a complex form and a seasonality can be overcome by means of the use of small window lengths. In general, there are some ad hoc rules-of-thumb to intuitively detect plausible groups of eigentriples:

- via a scree plot of the eigenvalues,
- w -correlations among possible groups,

Recently, (Mahmoudvand et al., 2013) suggested adapting SSA with optimal choices of parameters L and r based on the predictive forecast errors. In their study the time series are split up into two parts: training, and testing (for evaluation). They start with $L = 2$ ($2 \leq L \leq \frac{N}{2}$), and in the process consider all possible values of L and r ($1 \leq r \leq L - 1$) and singular values (for the selected L). Next the prediction Root Mean Square Error is determined as a loss function to minimize the forecast error. Ultimately the combination of L and r with the lowest loss function are captured, which represents the optimal SSA choices to decompose the series comprising of the validation set, and can be used for forecasting test values (Hassani et al., 2015). The main disadvantage of using this method is that the SSA parameters will be mostly dependent on the size of the training set, which is a trade-off between the amount of information one hopes to retain and the accuracy of the forecast. Also, if we were interested in capturing and analysing, say, only the seasonal components or the trend of the series, we would not be able to get the best possible decomposition and reconstruction for such purposes using this automated approach. If our objective is such, then it is more reliable to return to other approaches, which will enable one to analyse each paired eigenvector, and select those representing the seasonal components which are of interest to us.

Once an appropriate set of eigencomponents has been selected then MSSA can be used for forecasting the time series, therefore, satisfying a linear recurrence formula.

2.1.4 Linear Recurrent Formula (LRF)

If a time series evolves according to a state space model in an embedded dimension then its evolution in time can also be described by a linear recurrent formula (Golyandina et al., 2001) with coefficients $\{\phi\}_{j=1}^{L-1}$ such that:

$$y_t = \sum_{j=1}^{L-1} \phi_j y_{t-j}, \quad (2.4)$$

It means that SSA can essentially handle functions that are governed by an LRF and includes the broad class of functions that was proposed by (Buchstaber, 1994). In SSA the recurrent coefficients $\Phi = (\phi_{L-1}, \phi_{L-2}, \dots, \phi_1), j = 1, \dots, L-1$, are obtained as follows (See Section 4.2.1 for a derivation). First distinguish the first $L-1$ components of the eigenvector U_i , such that $U_i = [U_i^{\nabla L}, \pi_i]$ where $U_i^{\nabla L} = (u_{i,1}, \dots, u_{i,L-1})$ and π_i is the last component of U_i ; and define $\nu^2 = \sum_{i=1}^r \pi_i^2 < 1$. The recurrent coefficients are then estimated as:

$$\hat{\Phi} = (1 - \nu^2)^{-1} \sum_{i=1}^r \pi_i U_i^{\nabla L}. \quad (2.5)$$

In fact, the series governed by LRFs reveals natural recurrent continuation since each term of the series is equal to a linear combination of several preceding terms (Golyandina et al., 2001). A characteristic polynomial can also be assigned to the LRF which has $L-1$ roots. The theory of linear recurrent formulae and their associated characteristic polynomials goes back to (Gelfond, 1967). A more formal description on the concept of recurrent continuation in purely geometrical terms can be found in (Golyandina et al., 2001).

There exists different versions of forecasting methods for SSA, such as the minimal recurrent formula which removes the extraneous roots of the characteristic polynomial of the LRF (reduction method), and the nearest subspace approach when there is an approximate separability in which the selected linear space cannot be the trajectory space of the time series anymore (Golyandina et al., 2001). Specifically, in the presence of abrupt changes, the great number of the perturbed extraneous roots leads to a less precise forecast. On the other side, the conditions for approximate separability are typically asymptotic and need a comparatively large window length. Therefore, (Golyandina et al., 2001) suggests taking the smallest window length to provide sufficient separability. They pointed out that during a structural break a time series transitions from one *homogenous* state to another in a short period of time. Furthermore, while the time

series before and after the break may locally follow a LRF (or two separate LRF's) the combined time series may not. To detect such changes, they defined a heterogeneity function and an eigentriple rearrangement by measuring the discrepancy between two (or more) different states of a time series. They have also examined the root function of the characteristic polynomial of the LRF, which gives a dynamic description of the sequence of the homogeneous part of the series, which is affected by the eigentriple rearrangement and structural changes. As a result, the series stops continuing the original LRF, which means the LRF does not coincide with a recurrent continuation of the series before being perturbed. Thus, these changes in time series behaviour ought to be recognised by defining new initial conditions for the LRF. This is the core principle involved in this study which will be discussed later.

2.1.5 Extensions to basic SSA

MSSA is a direct extension of basic SSA to simultaneously analyse multiple time series. There are many different ways to modify and extend the basic SSA some of which are discussed below.

2.1.5.1 Vector SSA

The basic MSSA (or SSA) forecasting algorithm can be modified in several ways. For example, using Toeplitz SSA or SSA with centring (centre the trajectory matrix before calculating the SVD) for the time series with linear-like tendencies, and also for the stationary series. Perhaps one of the most important modifications of basic MSSA is the so-called MSSA vector forecasting algorithm (Golyandina et al., 2001).

Unlike the basic MSSA forecasting algorithm, which performs a straightforward recurrent continuation of the time series under study, the vector MSSA forecasting performs the continuation of the vectors in an r -dimensional space and only then returns to the time-series representation (Golyandina et al., 2001). The fundamental assumptions (embedding, SVD, grouping and diagonal averaging) of the vector forecasting algorithm is the same as the basic SSA forecasting algorithm. However, the key idea of vector forecasting is that the trajectory matrix of a time series can be continued by a sequence of vectors. In doing so, the following matrix can be formulated:

$$\mathbf{\Pi} = \mathbf{U}^\nabla \mathbf{U}^{\nabla T} + (1 - v^2) \Phi \Phi^T \quad (2.6)$$

where $\mathbf{U}^\nabla = [U_1^\nabla, \dots, U_r^\nabla]$. $\mathbf{\Pi}$ is the matrix of the linear operator that performs the orthogonal projection $\mathcal{R}^{L-1} \mapsto \mathcal{L}_r^\nabla$ where $\mathcal{L}_r^\nabla = \text{span}\{U_1^\nabla, \dots, U_r^\nabla\}$. Next, the following linear operator can be defined as: $\mathcal{P}^{(v)} : \mathcal{L}_r \mapsto \mathcal{R}^L$, where

$$\mathcal{P}^{(v)}Y = \begin{pmatrix} \mathbf{\Pi}Y_{\Delta} \\ \mathbf{\Phi}^T Y_{\Delta} \end{pmatrix}, \quad Y \in \mathfrak{L}_r \quad (2.7)$$

Y_{Δ} is a vector of last $L - 1$ components of Y . Consequently, vector forecasting can be broken into two steps:

1. Define the vector $Z_i^{(m)}$ ($m = 1, 2, \dots, M$) as follows:

$$Z_i^{(m)} = \begin{cases} \tilde{X}_i^{(m)} & \text{for } i = 1, \dots, K, \\ \mathcal{P}^{(v)}Z_{i-1}^{(m)} & \text{for } i = K + 1, \dots, K + h + L - 1, \end{cases} \quad (2.8)$$

where $\tilde{X}_i^{(m)}$ is the i^{th} reconstructed column of the trajectory matrix of the m^{th} series after grouping and eliminating noise components, and

2. Construct the matrix $\mathbf{Z}^{(m)} = [Z_1^{(m)}, \dots, Z_{K+h+L-1}^{(m)}]$ and then applying hankelization which results in a reconstructed time series $\hat{y}_1^{(m)}, \dots, \hat{y}_{N+h+L-1}^{(m)}$, such that $\hat{y}_{N+1}^{(m)}, \dots, \hat{y}_{N+h}^{(m)}$ indicates h^{th} terms of vector forecasting.

Although the fundamental assumptions for both algorithms are the same, there are practical differences in their forecasts. Both have two basic stages: diagonal averaging and continuation. In the basic SSA forecast algorithm, first diagonal averaging is used to determine the reconstructed series, and then continuation is applied using the LRF. While vector forecasting method uses these two stages in the reverse order. First, vector continuation in \mathfrak{L}_r is performed and then diagonal averaging yields the forecast values.

If there exists a recurrent continuation of the time series, then the results of forecasting for both vector and basic SSA coincide. On the other hand, in the case of approximate continuation the two forecasting algorithms usually differ and it is hard to compare their differences theoretically. According to a suggestion made by (Golyandina et al., 2001), the vector forecasting algorithm appears to be more *stable* and *conservative* than the basic algorithm especially for long tem forecasting. This was later confirmed in (Pepelyshev, 2010) who agreed with (Golyandina et al., 2001) based on a single application. Recurrent forecasting is simpler to interpret due to the description of LRFs in terms of the characteristic polynomials. Given the lack of a statistically reliable nature in the experiments underlying the aforementioned conclusions, it is not easy to conclude with absolute confidence as to which of the two approaches are best, or whether the best approach for a certain situation can be selected based on the structure of a given time series.

2.1.5.2 Bootstrap SSA

Forecasting methods can be assessed either by Monte Carlo simulations or via a bootstrap. However, Monte Carlo simulations can only be applied in situations when the

true model is known. In SSA, the true model for the signal is not known before filtering and reconstruction, and thus the bootstrap procedure is applied to obtain the statistical properties of the forecasts and construct interval estimates.

Assume that we have a time series which consists of two components $Y_N = \{y_t\}_{t=1}^N = S_N + E_N$ where S_N is the signal and E_N is the noise. Under a suitable choice of the window length L and the corresponding eigentriples, there is a representation $\hat{Y}_N = \hat{S}_N + \hat{E}_N$, where \hat{S}_N (the reconstructed series) approximates S_N and \hat{E}_N is a vector of residuals. Via bootstrapping (with replacement), we generate B independent *copies*, $E_{N,i}; i = 1, \dots, B$, of \hat{E}_N from the noise. We can then obtain B series $\hat{Y}_N = \hat{S}_N + \hat{E}_N$ and produce h step ahead forecasts \hat{S}_{N+h} in the same manner as in the Monte Carlo simulation. Average bootstrap forecasts can then be computed from the sample $\hat{S}_{N+h,i}; (1 \leq i \leq B)$ of these forecasts \hat{S}_{N+h} and be compared with the forecasting results obtained by the basic SSA.

The most widely used model for \hat{E}_N is the model of Gaussian white noise. The corresponding hypothesis can be checked with the help of the standard tests for randomness and normality (Golyandina et al., 2001). The bootstrap simulation uses the information obtained by the reconstructed series, which is different from the original signal. Therefore, the simulated confidence intervals are constructed for the entire series Y_N , while assuming that the series has the same structure in the future. It is shown by (Golyandina & Zhigljavsky, 2013) that these intervals obtained by the bootstrap simulation are rather stable as time moves on while the empirical ones are not. Thus, bootstrap prediction intervals can only be suitable for relatively short-term forecasting.

2.1.5.3 Monte Carlo SSA

Intuitively, basic SSA assumes the noise, E_N in the signal, plus the noise model as a *structureless* component of the resulting SSA decomposition. (Allen & Smith, 1996) shows that in some circumstances these noises can be considered as an AR(1) model, stochastic and coloured noise. This method is called Monte Carlo SSA, which can essentially improve the quality of the separation of signal from noise. To check the performance of separability some Monte Carlo simulations tests are applied (Allen & Smith, 1996; Golyandina & Zhigljavsky, 2013).

Monte Carlo SSA is also useful to test a null hypothesis that the oscillations are generated by an AR(1) process in paleoclimate time series (Yiou et al., 2000; Roth & Reijmer, 2005; Feliks et al., 2010). It has also been applied to study sea level cycles based on satellite data to check correct noise structure specification in the model prediction (Jevrejeva et al., 2006).

2.1.5.4 Spatial SSA

It is mentioned before that MSSA associates dynamical behaviour of multiple time series to a spatial temporal characteristic of their covariance matrix. Before exploring more about spatial SSA a brief introduction is provided on using spatial models for econometrics time series.

The origin of spatial econometrics is usually associated with the idea of spatial autocorrelation by (Cliff & Ord, 1972). Since then, the model has attracted a lot of attention. Positive spatial autocorrelation appears between regions, in close proximity to one another, with similar values which are clustered together in space. On the contrary, negative spatial autocorrelation occurs when dissimilar values for a variable are clustered together in space. According to (Trendle, 2002) the space is introduced into the estimated procedure of the spatial weights matrix, W , and measures the proximity of regional states. In principle, the major contribution made by the spatial weight matrix is its connection between a specified variable and the related variable in another spatial location (Anselin, 2013).

Based on the previous studies a country or a region's growth can effectively depend on the growth of other countries or regions which is called spatial dependency. Literature on spatial time series forecasting and labour markets has focused almost exclusively on parametric time series models, in both its linear and nonlinear models. A comprehensive study on different modelling approaches for forecasting can be found in (Floros, 2005). This study compares the performance of twenty-three models, including different types of ARMA, ARIMA, GARCH, TVAR, STAR etc. Their findings on the UK labour market show that there is a close relationship between labour market conditions and forecasting records. Later on, (Schanne et al., 2010) showed that a spatial GVAR model can be a better alternative, or a complementary approach to those commonly used models for regional forecasting, in which regional interdependencies are not taken into account. For example, (Schanne et al., 2010) model regional unemployment in German labour-market districts, and argue in favour of allowing for the possibility that economic development in some labour markets might affect neighbouring or close regional markets. In these studies, the preliminary assumption is that the given time series is stationary, which can be indeed restrictive and not viable in reality. Despite widespread recognition of their importance, we found a notable lack of research regarding the usage of nonparametric models on spatial time series forecasting and labour markets.

(Awichi & Müller, 2013) use SSA with a spatial weighted average, to forecast a set of rainfall recordings from several locations in Upper Austria. Their results show that MSSA is slightly better than the univariate one but only for in-sample forecasting. It is also argued that the results are considerably sensitive to the selection of window length and the number of eigenvalues (Awichi & Müller, 2013). Therefore, more studies are

necessary to further develop MSSA, whilst allowing for the presence of spatial interdependency among the variables of interest without restricting its forecasting performance. Indeed, in most practical situations it is natural to suppose them to be functions of the underlying states, and therefore, be liable to change as location changes. This recursive specification is the inherent core in a Bayesian time series, which creates model extrapolation by developing the state via the state evolution equation into the future. The last section of this chapter gives a brief introduction on Bayesian forecasting models and its related works.

2.2 Bayesian Forecasting

Bayesian methods differ from classical methods in that every unknown in a system may be assigned a distribution as opposed to having some unknown but fixed value. These distributions are then updated as observations become available moving us from a prior state to an a posteriori state. The original Kalman filtering algorithm is one early example of a Bayesian approach to time series forecasting and since then several variants have emerged such as the extended Kalman filter (Gelb, 1974; Meinhold & Singpurwalla, 1983), the unscented Kalman filter (Julier & Uhlmann, 1997) and most recently particle filtering (Del Moral, 1996).

Using a Bayesian framework in analysing time series and forecasting was first introduced by (Stevens & Harrison, 1971; Harrison & Stevens, 1976) as an effective model to distinguish transient changes from sudden real events. An extensive number of concise studies and precise investigations shows that Bayesian forecasting has considerable advantages over its classical counterparts such as ARIMA, ARFIMA, exponential smoothing and also nonlinear models like ARCH (see for example (Pole, Andy and West, Mike and Harrison, 1994; Steel, 2001; Granger, Clive William John and Timmermann, 2006; West, 2013)). There are many reasons why a Bayesian approach to modelling has attracted such popularity such as (Stevens & Harrison, 1973):

- Comprehensive model coverage. Almost all classical linear forecasting approaches like Box-Jenkins and exponential smoothing models can be formed in the Bayesian format.
- Assuming model's coefficients as stochastic variables (for example, to cater for drifts) to attain their interaction with the independent variable.
- To forecasting with little or no direct data (when a prior exists).
- Being of great value for a mixed model specifically at times of significant changes which allow ultra-stable control.

- Establishing a natural means for system communication by offering dialogue between the formal statistical model and the human controller.
- Dealing with nonlinear models and unequally spaced series.
- Expandable to multivariate forecasting and those subjects to constraints on totals or subtotals.

Alternatively, pointing out the advantage of fitting a recurrence structure to an estimation process does not lead this research to ignore the importance of traditional data analysis methods such as cross-correlation, spectral analysis and decomposition analysis. These methods are still reliable and valuable for certain occasions (Tong, 1990).

Note that the Bayesian approach to time series forecasting is a natural bridge between the theoretical work done in time series and practical forecasting experience. During the last couple of decades, methodological and practical applications of Bayesian time series analysis have been greatly extended to non-Gaussian, non-linear and more complicated conditionally linear models (Steel, 2001).

In essence, there are different ways to classify time series analysis techniques: “parametric and non-parametric” or “linear and non-linear” or “univariate and multivariate” including various combinations. For example, bilinear models, threshold autoregressive models, exponential autoregressive models, non-linear moving average models and doubly stochastic models are non-linear parametric models (Tong, 1990; Teräsvirta et al., 2008), while Fourier transformation models and projection pursuit models are non-parametric non-linear models (Fan & Yao, 2003; Gao, 2007) (and also a mixture of parametric and non-parametric forms which are called semi-parametric (Granger & Teräsvirta, 1993; Tong, 1990)). Such studies have also focused on using non-parametric methods for non-linear and non-stationary time series, and more examples can be found in (Phillips, 1998; Karlsen, Hans Arnfinn and Tjostheim, 2001) on non-parametric AR models, (Bandi, Federico M and Phillips, 2003) on non-parametric estimation of non-stationary scalar diffusion process, (Gao & Tong, 2004) on cross-validation-based model selection procedure for the simultaneous choice of time series components, (Wang, Qiy-ing and Phillips, 2009) on kernel estimation of structural non-parametric augmented regression for non-stationary time series, and (Karlsen et al., 2007) on structural non-parametric co-integrating procedure to address endogeneity.

2.2.1 Linear vs Nonlinear

One of the most well-known classes of forecasting models are linear models. They can obtain accurate but not error free prediction, which mainly require a specification that should match the underlying data generation process (Geweke & Whiteman, 2006). For example, it is assumed that data points taken over time may have an internal structure

(such as autocorrelation, trend, seasonal or cyclical fluctuation) that should be accounted for (Davis et al., 2006). They also embody some restrictive parametric conditions such as normality, stationarity and linearity before processing the data (Brockwell & Davis, 2002). Therefore, these methods perform weakly on identifying complex attributes which occur due to the aim of characterising all time series perceptions, the requirement of stationary, residual normality and independency (Dickey et al., 1991; Harrison, 1967). The forecast horizon is also a factor in determining the correct model specification; with some models being particularly suited to predictions in a longer term horizon (Chatfield, 2013).

Furthermore, Autoregressive (AR) based models are basically retrogressive looking. As such, they are generally poor at predicting turning points, unless the turning point represents a return to long-run horizons (Franses, 1998). Moreover, linear models can only interpret regular behaviour like exponentially decaying (or growing) or a periodically oscillating structure, thus those intrinsic dynamic and irregular behaviour of the series have to be attributed to some sort of non-linear paradigms (Kantz & Schreiber, 2004). There is, indeed, an interesting type of duality between the concepts of non-stationarity and non-linearity. A common method for managing both concepts is to separate the parameter space into a large number of small segments, and consider the process as locally stationary (or linear) inside each segment. A formal description of this method prompts the development of a non-stationary process and non-linear models in respect to evolutionary time dependent spectra in (Priestley, 1965) and later on used by (Huang et al., 1998; Zhong, 2006).

Providing such models with more flexibility may result in over-fitting in a sample. Therefore, it is recommended by (Teräsvirta et al., 1993) to first test linearity against non-linearity and then only if linearity is rejected then build a non-linear model. There are a wide range of tests to detect linearity, such as a non-parametric non-linear test proposed by Hjellvik and Tjstheim (1995), in which it may assist the practitioner in deciding whether to apply a non-linear model rather than a linear one. The test calculates the discrepancy between the best linear and the best non-linear predictor (more can be found in (Hardle et al., 1996; TeIjdsvirta, 1990; Lee, T.-H., H. White & Granger, 1993)). (Gao & Tong, 2004) claims that semi-parametric methods can provide a model with better predictive power than non-parametric methods but this is highly data set dependent. Later on, (Gao et al., 2009) proposed a (non-parametric kernel) test for non-linearity and non-stationarity which can be applied to a class of non-linear autoregressive mean models. The test process is mainly associated with its parametric version test to include non-stationary, also known as the Dickey-Fuller test. Among all these tests only a few of them can handle various situations simultaneously (Teräsvirta et al., 2008), squashing function based models (except for data generated by bilinear models) and Lagrange multiplier (LM)-type test (Zivot & Wang, 2006).

2.2.2 Nonparametric vs parametric

Basically, building a nonparametric model can decrease the subjectivity of selecting a particular type of a parametric model before exploring the given dataset (Hardle et al., 1996). Even though complexity of nonparametric models brings more difficulties into the model (such as choosing smoothing parameters), and sometime causes poor performance in high dimensional cases. Hence, the nonparametric approach often makes a general assumption by giving priority to choose an appropriate lower dimensional model, and then picks up the most informative features of interest through the datasets. Unlike the parametric version of it, which select a particular structure from a parametric class by estimating a fixed finite number of parameters, a non-parametric model estimates covariance or the spectrum of the series without assuming that the series has any particular structure, and will update the approximation precision when new information becomes available (Hardle et al., 1996). In line with prediction and forecasting, these non-parametric methods mainly focus on conditional means (if a point forecast is desired), conditional variances (if interval forecast of future volatility is desired) or complete conditional densities in some periods (if higher order moments of a series are potentially important), given the past of the process (Hardle et al., 1996). The following models are suggested to approximate a conditional mean: index models (Bierens, 1988), non-linear additive AR models (Hastie, Trevor J and Tibshirani, 1990), adaptive spline threshold AR model (Lewis, Peter AW and Stevens, 1991) and functional coefficient AR model (Hardle et al., 1996), and conditional variance: density estimation with correlated observations using non-parametric kernel methods (Tjøstheim & Auestad, 1994), local polynomial regression (Hardle, Wolfgang and Tsybakov, Alexandre and Yang, 1998) and a Hermite expansion approach (Gallant, A. Ronald & Tauchen., 1989).

One important key point, which allows nonparametric models to be functional and practical, is that they should be able to establish a direct relation between new economic theories and both linear and non-linear econometrics models (Pesaran & Potter, 1992). Therefore, it completely implies the necessity of developing potential models based on non-linear structural models in both theoretical and empirical research. It should be emphasised, however, that the lack of structural models is one of the most crucial facts about many nonlinear time series models. According to (Pesaran & Potter, 1992), there are two separate but closely attached reasons for that: First, the absence of a closed form formula and the qualitative nature of many of the results in non-linear dynamic or chaos theory, which means that there is no clear initial point. Second, however, one might believe that linear models aggregate to produce linear models, and it is not possible to believe that nonlinear models aggregate to nonlinear models in the same class (Pesaran & Potter, 1992). Hence, most theoretical and practical studies on non-linear time series modelling focused mainly on developing non-linear structural models. One of the most commonly used approaches is using Bayesian inference for time series modelling.

It seems to be widely accepted that the relationship between economic time series is mostly non-linear (Priestley, 1980; Franses, 1998; Zivot & Wang, 2006). In other words, most economic time series contain atypical observations, clustering of outliers and non-linearities, (Franses, 1998). The key examples of these sort of series are industrial production indices and unemployment rates (Gupta & Kabundi, 2011). These series display asymmetric patterns or erratic behaviours accompanied with high frequency when the economy is overheating compared to when it is underperforming (Priestley, 1980). Therefore, those non-linear and asymmetric behaviours must be included in modelling the series and forecasting (Haggan et al., 1984).

2.2.3 State Space Models

For analysing nonlinear systems, both deterministic and stochastic dynamical systems, the general framework of state space models is very useful. State space models enable the modeling of a variety of unobservable but interpretable dynamic components by applying Kalman filtering to estimate them in such a way that they can handle shifts, structural breaks and time-varying parameters of some static models. A state space model can also handle irregular or unequally spaced data in two different ways. First by defining the model, which varies over different time intervals, and second by forming a regular interval while treating empty intervals as missing values (Harrison & West, 1987). In general, there are sufficient benefits of using state space models which is listed below:

- Handling missing values naturally
- Incorporating explanatory variables into the model without difficulty
- Allowing regression coefficients to vary stochastically over time
- Assuming no additional theory for forecasting subsequent to all that is required to extend the Kalman filter forward into the future. To find more about them see (Harrison & West, 1987; Durbin & Koopman, 2012).

There is also a comparatively long history of using state space models in econometrics, for example a well-known reference goes back to (Harrison, 1967; Harvey, 1984), and a more precise exploration is provided by (West & Harrison, 1997). Later on (Koop, 2003) provide a comprehensive review of the application of Bayesian analysis to state space models.

These state space relations can be often modelled by a set of differential equations. The mathematical equations depicting the system can be deterministic or stochastic, depending on the way the differential equation is constructed, and the dynamical system

might be linear or non-linear. The main idea of state space models is that a time series y_t is generated by an observation or measurement equation:

$$y_t = H_t \theta_t + \epsilon_t \quad (2.9)$$

while the transition between the state vectors, θ_t (determining levels, trends or seasonality), can be dynamically modelled by the following system known as the transition equation:

$$\theta_t = F_{t-1} \theta_{t-1} + W_t \quad (2.10)$$

H_t is an inherent characteristic of the system conducting its motion and F_{t-1} is a state transition matrix and is typically characterised in a block diagonal form and opens to modification over time (West, 2013). Depending on the application area, errors ϵ_t and W_t can be considered in different forms (They are usually considered as mutually independent variables with Gaussian distribution). Technically there is no need to assume a Gaussian distribution for errors, however, defining a prior distribution is a necessity to take the initial values of the state vectors (Steel, 2001). A simple case of a state space model is the conventional linear model mostly used by classical statisticians and can be obtained by replacing θ_t with θ and considering W_t as an independent variable of time. Alternative special cases can be set by choosing different types of F_t and H_t in both measurement equation and transition equation (Harrison & West, 1987). The most commonly used models are considered below:

- A Random walk driven by noise
- Local trend/polynomial dynamic linear models
- Dynamic regression
- Seasonal dynamic linear models
- Autoregressive and time varying dynamic linear models

Another way of categorising a linear state space model or a structural time series is modelling each single component differently. Different time series components, like trend, seasonal, cyclical and irregular variability can be formed individually before being formulated in the space state model. On the other hand, regarding which component is added to the model, there can be a broad variation of state space modelling such as:

- Trend component: Sum up the model with a slope term which is generated by random walk and called a local linear trend model.
- Seasonal component:
 - Time domain effect

- Quasi-Random walk, (Harrison & Stevens, 1976)
 - Trigonometric effect
 - Time-Varying stochastic seasonal effect, (Young, Peter Colin and Ng, Cho Nam and Lane, Kevin and Parker, 1991)
 - Trigonometric effect with Quasi-Random walk model (frequency domain model)
 - Trigonometric effect with stochastic model
- Cyclic component
 - Explanatory variables and intervention effects

Data from various information sources can be incorporated into the model for prediction and estimation. (Harvey, 1984) defined a unified framework for univariate time series regarding trend, seasonality and irregular components and reviewed an extensive variety of extrapolative forecasting techniques to establish a relationship between them. In his study he came to the conclusion that Kalman filtering can be the integrated treatment of a variety of forecasting methods (Harvey, 1984). Basically, a Kalman filter is recommended because of its ability to establish a recursive relationship between the model parameters for either updating or revising (which is considered as a filtering process) (Turner, 2011).

2.2.3.1 Kalman Filtering

The early stage of state space models was established by control engineering rather than by statisticians, as found in the publication by (Kalman, 1960). In his study, Kalman not only proposed that a wide range of current models can be formed by a state space model, but he also proves that the computation necessity for the practical application of state space models could be set up in sequential form mainly because of their Markovian characteristic. The principle aim of this method is updating the observation equation at each time interval by the parameters, which evolve over time and which are not directly observable. Provided that a model can be formed as a state space model, a wide range of significant and interpretable statistical analysis can be made immediately available. Henceforth, it is not just the optimal estimation of future states that can be accomplished using the Kalman filter, additionally, the foundation of a parameter estimation can be also be done by the Kalman filter on the grounds that it is able to specify the likelihood function in the matter of one step ahead forecast errors (Harvey, 1984).

The Kalman Filter is an optimal estimator for linear system models and consists of two steps: prediction and updating. In the first step, it permits evaluation and estimation to be finished inside a well-organised system and, besides, it permits forecast mean square errors to be measured.

Theorem 2.3. *The Kalman Filter. The Bayesian filtering equations for the linear filtering model 2.9 and 2.10 can be evaluated in closed form and the resulting distributions are Gaussian:*

$$\begin{aligned} p(\theta_k|y_{1:k-1}) &= N(\theta_k|m_k^-, P_k^-), \\ p(\theta_k|y_{1:k}) &= N(\theta_k|m_k, P_k), \\ p(y_k|y_{1:k-1}) &= N(y_k|H_k m_k^-, S_k). \end{aligned} \quad (2.11)$$

The parameters of the distribution above can be computed with the following Kalman filter prediction and update steps.

1. The prediction step is

$$\begin{aligned} m_k^- &= F_{k-1} m_{k-1}, \\ P_k^- &= F_{k-1} P_{k-1} F_{k-1}^T + Q_{k-1}. \end{aligned} \quad (2.12)$$

where $\epsilon_t \sim N(0, Q_{k-1})$, $W_t \sim N(0, R_k)$ and $\theta_0 \sim N(m_0, P_0)$.

2. The update step k is

$$\begin{aligned} v_k &= y_k - H_k m_k^-, \\ S_k &= H_k P_k^- H_k^T + R_k, \\ K_k &= P_k^- H_k^T S_k^{-1}, \\ m_k &= m_k^- + K_k v_k, \\ P_k &= P_k^- - K_k S_k K_k^T. \end{aligned} \quad (2.13)$$

The recursion is started from the prior mean m_0 and covariance P_0 .

This technique relies substantially on an inherent inference in which more weight ought to be given to the latest observation while repeating the procedure systematically. The theoretical implementation of this method is well studied by (Grewal, 2011; Grewal et al., 2007; Sarkka, 2013).

A nonlinear version of the Kalman filter, which describes how non-linear dynamics may be approximated by a local linearisation, is called an Extended Kalman Filter (EKF). One of the most common versions of EKF are the Taylor based EKF. It uses a Taylor series approximation about the transition function (no longer a matrix).

Theorem 2.4. *The Extended Kalman Filter. The prediction and update steps of the first order additive noise extended Kalman filter are:*

1. Prediction:

$$\begin{aligned} m_k^- &= F(m_{k-1}), \\ P_k^- &= F_\theta(m_{k-1})P_{k-1}F_\theta^T(m_{k-1}) + Q_{k-1}. \end{aligned} \quad (2.14)$$

2. Update:

$$\begin{aligned} v_k &= y_k - h(m_k^-), \\ S_k &= H_\theta(m_k^-)P_k^-H_\theta^T(m_k^-) + R_k, \\ K_k &= P_k^-H_\theta^T(m_k^-)S_k^{-1}, \\ m_k &= m_k^- + K_k v_k, \\ P_k &= P_k^- - K_k S_k K_k^T. \end{aligned} \quad (2.15)$$

where F_θ is the Jacobian matrix of F .

The main requirement of the EKF algorithm is that both the observation equation and the transition equation ought to be differentiable. Therefore, in the so-called second order EKF second order terms of Taylor series expansions is taken for tracking a nonlinearity. There are other types of EKF such as the statistically linearised filtering (Gelb, 1974) (replace the first order Taylor series expansion with a statistical linearisation), the Unscented Kalman filter (Julier et al., 1995; Wan & Van Der Merwe, 2000) (directly approximates the mean and covariance of the target distribution instead of trying to approximate nonlinear functions), Gaussian filtering (Maybeck, 1982) (use Gaussian assumed density approximation and approximate its mean and covariance via moment matching), the Gauss-Hermit Kalman filter (Ito & Xiong, 2000) (replace Gaussian integrals in the Gaussian filter algorithm with Gauss-Hermit algorithm) and others (Ito & Xiong, 2000; Sarkka, 2013).

2.2.3.2 State Dependent Models

As mentioned earlier, nonlinear models can be categorised in different ways. The most commonly used nonlinear models are Bilinear models (Mohler, 1973), Threshold autoregressive models (Tong & Lim, 1980) and Exponential autoregressive models (Haggan & Ozaki, 1981). For example in bilinear models the parameters are a linear function of $\epsilon_{t-i}y_{t-i}$, in Threshold AR models parameters are a step-function of y_{t-d} , and for exponential AR models they are an exponential function of y_{t-1}^2 . Each of these classes is a particular type of non-linearity, and therefore, this makes it hard to decide which one is the most appropriate model for a given set of data. According to (Priestley, 1980) a nonlinear model should be able to capture a better fit to the data with more flexibility, as well as uncovering interesting patterns such as limit cycles, amplitude-dependent frequency and jump phenomena, which can never be determined easily by linear models.

Therefore, he proposed the State Dependent Model as a general class of non-linear models, because of its greater flexibility to cover non-linear time series models, as well as standard linear models as special cases. It uses a sequential type of recursive algorithm to distinguish state dependent models, and determines their applications in forecasting and detecting non-linearity.

Practically, using state dependent models is advantageous in two aspects. First they can be used straightforwardly in line with prediction and forecasting, and secondly, they can provide an overview of the inherent nonlinear characteristics in a given series to check the credibility of the fitted model (like bilinear, threshold autoregressive and exponential autoregressive), as they can be fitted to the data with no particular prior information about the type of non-linearity. A locally linear ARMA model, for example, may be expressed as:

$$y_t = \mu(\theta_{t-1}) + \sum_{u=1}^k \Phi_u(\theta_{t-1})y_{t-u} + \sum_{u=1}^l \Psi_u(\theta_{t-1})y_{t-u} + \epsilon_t, \quad (2.16)$$

where $\{\Phi_u\}$ are the AR coefficients, $\{\Psi_u\}$ are the MA coefficients, μ is a local mean, and note that all of the parameters are dependent on the state vector of the model at time $t - 1$.

Selecting different forms for μ , Φ and Ψ leads to the fact that the SDM includes, as special cases, linear models (ARMA), bilinear models, exponential autoregressive models and threshold autoregressive models. For example, by taking μ , Φ and Ψ as constants and independent of the state vector the (2.16) reduces to an ARMA model (For further detail see (Priestley, 1980)).

More rigorously, the state observation and transition equations (2.9) and (2.10) are basically determined by the potentially time varying quadruplets $\{F_t, G_t, \epsilon_t, W_t\}$ and time varying state vector θ_t , which makes them versatile to evolving circumstances. In state dependent models the transition matrix F_t is assumed to be tied down firmly to the state of the process at time $t - 1$. In other words, the coefficients can be evolving based on both time and the state of the process, and be automatically adjustable to shifts and structural breaks. Hence, μ , Φ and Ψ are considered as analytic functions of θ_t which change over time. A non-linear/non-stationary part of the model might be related to the dependence of Φ_u and Ψ_u on both θ_t and t (Priestley, 1981). To conclude what have mentioned before, the SDM model can be re-written as follows:

$$\begin{aligned} y_t &= H_t \theta_t + \epsilon_t \\ \theta_t &= F_{t-1}(\theta_{t-1}) \theta_{t-1} + W_t \end{aligned} \quad (2.17)$$

In SDM, the practical strategies for non-linear models are found, in essence, on splitting up the state space into a large number of small segments, and considering the process

as locally linear inside each segment. To paraphrase Priestley, by saying that the state dependent model is formed by *bending the linear model*, it means one is taking a local linear model at time t and bending it, we can obtain the best fit to the next observation, y_{t+1} . On the other hand, local can also imply small departures from the current state of the process, and in this case the coefficients would be time-dependent rather than state-dependent, and in parallel can describe non-stationarity rather than non-linearity (Priestley, 1981). Considering $\{\phi_u^{(i)}\}$ as a slowly changing function of $y_t^{(i)}$ this results in updating equations for the coefficients:

$$\phi_{u,t}^{(i)} = \phi_{u,t-1}^{(i)} + \Delta y_{t-u}^{(i)} \gamma_u^{(t)}, \quad u = 1, \dots, L-1, \quad (2.18)$$

where $\Delta y_{t-u}^{(i)} = y_{t-u}^{(i)} - y_{t-(u-d)}^{(i)}$ and γ_u is a *gradient*. The gradient parameters $\gamma_1^{(t)}, \dots, \gamma_{L-1}^{(t)}$ are unknowns, and must be estimated at time t . The basic strategy is to allow these parameters wander in the form of random walks. The random walk model for the gradient parameters may be expressed in matrix form as:

$$B_{t+1} = B_t + V_{t+1}, \quad (2.19)$$

where $B_t = (\gamma_1^{(t)}, \dots, \gamma_{L-1}^{(t)})$ and V_t is a sequence of independent matrix-valued random variables such that $V_t \sim N(0, \Sigma_V)$. The estimation procedure then determines, for each t , those values of B_t , which seeks to minimise the discrepancy between the observed value of y_{t+1} and its predictor \hat{y}_{t+1} , computed from the model fitted at time t . The algorithm is thus sequential in nature and resembles the procedures used in the EKF algorithm.

In fact, the relative magnitude of $\|\Sigma_\nu\|$, the variance-covariance matrix of V_t to σ_ϵ^2 should determine the sensitivity of the algorithm to changes in B_t . The choice of Σ_ν depends on the assumed smoothness of the model parameter as a function of y_t . If $\|\Sigma_\nu\|$ is small in comparison with σ_ϵ^2 then \hat{B}_t should not change rapidly but be effectively constant over time. On the other hand, if $\|\Sigma_\nu\|$ is large in comparison with σ_ϵ^2 then \hat{B}_t will change rapidly in order to ensure a good fit for each new observation. Priestley (Priestley, 1981) began the algorithm with an ARMA model to estimate the initial values like μ, Φ_u, Ψ_u and σ_ϵ^2 and consequently the values of Σ_ν will depend on how quickly the practitioner thinks the B_t are changing, i.e. Σ_ν depends on our assumed smoothness of the B_t . Selecting Σ_ν in which its diagonal elements are related to $\hat{\sigma}_\epsilon^2$, endow the updated stage with more flexibility in terms of changing of the model parameters. The diagonal elements of Σ_ν are set equal to $\hat{\sigma}_\epsilon^2$ multiplied by some constant α called the smoothing factor, and the off-diagonal elements are set equal to zero. However, if the elements of Σ_ν are set too large, the estimated parameters become unstable, but if the elements of Σ_ν are made too small, it is difficult to detect the non-linearity present in the data, since the procedure is then virtually equivalent to a recursive fitting of a linear model. The best procedure in practice appears to be to reduce the magnitude of

the smoothing factor until the parameters show stable behaviour. If the parameters still appear to be far from smooth, the smoothing factor may be reduced further. In addition, the parameters may be smoothed by a multidimensional form of the non-parametric function fitting technique (see for example (Priestley & Chao, 1972)). Having carried out this procedure, it is hoped that the resulting parameter surfaces give a clearer idea of the type(s) of non-linearity present in the model.

(Priestley, 1981) shows that an SDM is able to detect the linear part of a bilinear model more precisely than its bilinear part, as one would expect, because of the inherent difficulty of bilinear estimation. They also showed empirically that the SDM algorithm is not sufficiently accurate for linear threshold data generating processes, due to the model being unable to recognise an abrupt change in the coefficients. However, it can identify smooth changes of the coefficients of a non-linear data generating process. As a consequence, the algorithm requires more development and justification especially for linear threshold models to address the issue and make it flexible in detection of changes regardless of the fact of how sharp they are.

A smooth change of the gradient over time is yet another basic assumption that underlies the SDM algorithm. In the SDM algorithm the gradient parameters are assumed to follow a random walk, and a smoothing factor is defined to allow more flexibility in the variance of the innovations. The study by (Haggan et al., 1984) compares the actual behaviour of the gradient parameters and their estimates for linear threshold AR models, which display more variability, and hence the algorithm can be improved by considering the fact that the behaviour of the gradient parameters can be modelled by some stationary time series models rather than by a random walk. This form of behaviour can also be seen in other models like the bilinear, exponential AR and non-linear threshold models, and the smooth-transition autoregressive (STAR) models of (Chan & Tong, 1986) and (Granger et al., 1993).

The SDM has been recognised as a viable algorithm for modelling non-linear time series. But of course further improvement and study is needed in terms of modelling gradient parameters to enhance the accuracy and efficiency of the algorithm, especially for threshold models.

2.2.4 Bayesian SSA

A combination of SSA and Bayesian modelling is proposed by (Holmström & Launonen, 2013). They argue that there exists various source of uncertainty inside the time series, and thus it is hard to decide if the extracted signal includes true features of an underlying phenomena or just artifacts of noise (Holmström & Launonen, 2013). Thereupon, they

suggest that a Bayesian verification⁷ is expected to affirm the validity of the oscillatory periods found by SSA.

At that point, to discover an interesting feature of an extracted signal underlying the time series, such as trends and periodicities, they project the posterior from samples taken from the distribution of the eigenspace generated through SSA. Bayesian inferences are then made upon the credible features in these projections. The slopes of the projected sample can then be used to examine whether the extracted underlying component is credibly presented in the time series. This approach is called Posterior SSA (PSSA) which basically borrows the idea of the BSiZer methodology, scale space technique that uses smoothing to reveal interesting features, as introduced by (Erasto & Holmstrom, 2005).

To best of our knowledge, PSSA is the only work that has conducted Bayesian modelling on SSA. Therefore, more studies and research are needed to further develop MSSA (or SSA) in other aspects as well, such as forecasting. In the next chapter, this research will report on our exploratory analysis about the core of SSA forecasting, the LRF coefficients, and thereafter it will show how the accuracy of a forecast can be improved by modifying the coefficients via state dependent models.

⁷In the sense that the extracted modes are statistically significant.

Chapter 3

Preliminary analysis

3.1 Introduction

As discussed in Section 2.1 the primary aim of SSA is to decompose the original time series into the sum of a small number of independent and interpretable components such as a trend, oscillatory component(s) and noise. Therefore, SSA enables us, upon reconstruction of the series under study, to produce forecasts for either the individual components of the series and/or the reconstructed series itself. This is useful if ones want to make predictions about, for example, the deterministic/trending component of the series without taking into account the variability due to other sources (Hassani & Zhigljavsky, 2009).

The core of SSA forecasting lies with the recurrent coefficients of LRF (Formula 2.4 in Section 2.1.4). In this Chapter we perform a preliminary analysis of the LRF coefficients.¹ We begin our discussion by looking briefly at some standard techniques to statistically approximate the distribution of these coefficients for further analysis. Thus, bootstrapping is employed, an important nonparametric tool, for assessing and improving the properties of coefficients and their forecasts. As a resampling method, a bootstrap approximates the shape of the sampling distribution by simulating replicate experiments on the basis of the data and calculates a bootstrap estimate of the statistic.

As explained in Section 2.1.5.2, a bootstrap SSA forecast is proposed by (Golyandina & Zhigljavsky, 2013). However, there are some issues regarding the bootstrap SSA forecast. This approach is very sensitive to noise, in addition, the level of uncertainty can increase rapidly with the forecast horizon. Our preliminary analysis suggests that a bootstrap can be used not only for replicating forecasts but also for resampling the LRF coefficients. In this way, we will have some empirical information about the coefficients which allows us to statistically approximate their underlying characteristics. Furthermore, it

¹The results described in this chapter have been published in *International Journal of Energy and Statistics* (Rahmani, 2014).

will be shown that using a bootstrap of the LRF coefficients may improve forecasting accuracy. Rather than producing a point estimate forecast, it can likewise be utilised to approximate Forecast Intervals (FI) for assessing future uncertainty, comparing various forecast results more thoroughly and exploring different scenarios based on different assumptions more carefully.

The efficiency of the *bootstrapped LRF coefficients* is tested by using both synthetic data and some real data sets for different levels of noise and window lengths. In particular, the results would seem to suggest that the average bootstrap LRF coefficient results in a more reliable forecast for various window lengths and at different noise levels. In addition, the empirical results affirm that updating coefficients brings about narrower and smoother boundaries in comparison with the bootstrap SSA forecast.

3.2 Bootstrap LRF

Theoretical approaches to extract information about the LRF coefficients would require necessary assumptions such as identification of their correct model or knowing the generating process of the time series. In contrast, a bootstrap approach does not require any assumption about the validity of the model, the form of their generating process or the form of the distribution of the forecast errors. The only requirement for bootstrapping is that data is independently sampled from a single source distribution. As stated in Section 2.1.5.2, bootstrap SSA (Golyandina et al., 2001) uses a bootstrap in the residuals E_N , due to the fact that from the white noise extracted by SSA it is possible to ensure independence, the required condition for application of the method.

In order to forecast at a horizon, h , using bootstrap SSA, B (replicate) forecasted values are generated for each time horizon which can be used to calculate lower and upper percentiles of each observation to establish a boundary for the forecast. We also concluded that it might be useful to use other related information provided by bootstrap SSA. It was then observed that per iteration of bootstrapping SSA provides information regarding the underlying coefficients of the LRF which can be useful for further analysis. Thereupon, the bootstrap LRF coefficients were utilised to assess the properties of coefficients and improve their forecasts.

Of particular importance, bootstrap sampling provides some information about the coefficients Φ_i in each replication, B , which are used here to build a smoother forecast interval and as a result a point forecast. For this to happen, we simply consider the mean of the coefficients, $\bar{\Phi} = \frac{1}{B} \sum_{i=1}^B \Phi_i$ as an unbiased estimator to further analyse their distribution, forecasts and forecast intervals (which are called $FI_{\bar{\Phi}}$). The following section presents empirical results to gain a better understanding of the key role of the LRF coefficients in both synthetic and real time series.

3.3 Empirical Results

As discussed earlier, we aim to examine the performance of the bootstrap LRF coefficients in two aspects. Firstly, we want to see how much the accuracy of the forecasts can be improved by further investigation of the coefficients of the LRF (tested by prediction root mean square error (RMSE)). It is therefore of interest to measure the uncertainty level of the forecast intervals provided by these coefficients. Secondly, we explain how this uncertainty affects the behaviour of these coefficients.

One of the most important factors to evaluate the level of uncertainty of forecast limits is the criteria to measure what percentage of data is included in the predicted interval. The significance level of a forecast interval is usually denoted as $\alpha = 5\%$ to evaluate the *coverage percentage* of boundaries. A high percentage of coverage indicates that the provided boundary covers most of the signal (Christoffersen, 1998). It is necessary to take into account that wide boundaries also imply a high level of uncertainty. Therefore, the *asymmetry ratio* of a forecast interval is also calculated as another criteria to evaluate the reliability of forecast intervals (Christoffersen, 1998). The asymmetry ratio (ASR) of an FI is defined as:

$$ASR = \frac{Y - L_{FI}}{U_{FI} - L_{FI}} \quad (3.1)$$

where U_{FI} is the upper bound of FI, L_{FI} is the lower bound of the FI, and Y is the original time series (O'Connor et al., 2001). If the asymmetry ratio is approximately 0.5, it shows a symmetric interval for a forecast. If the value is greater than 0.5, the acquired lower bound tends to be further away from the upper bound (O'Connor et al., 2001).

The following section evaluates the performance of the bootstrap LRF coefficients for a synthetic time series with respect to bootstrap SSA. In addition, the effect of the signal to noise ratio and window length are examined.

3.3.1 Synthetic time series

We begin our experiment by using a simple time series with a deterministic trend. Define a time series, Y_N as $Y_N = \exp(\nu t) + E_N; (t = 1, \dots, N)$, where E_N is drawn from a normal distribution. When $L = 2$ and $r = 1$, based on the LRF, a one-step ahead forecast can be expressed as in Equation (2.4) as $\hat{Y}_{N+1} = \phi_1 Y_N$ where ϕ is obtained from the eigenvector of the extracted signal (Section 2.1.4). We expect that a different level of noise may result in a different ϕ_1 in each step and consequently different forecasts. Thus, the analysis considers different ranges of signal to noise ratio, from 4 to 15. In total 200 sample paths are generated each with 1000 steps. The first 100 samples are considered as in-sample and the rest as out-of-sample.

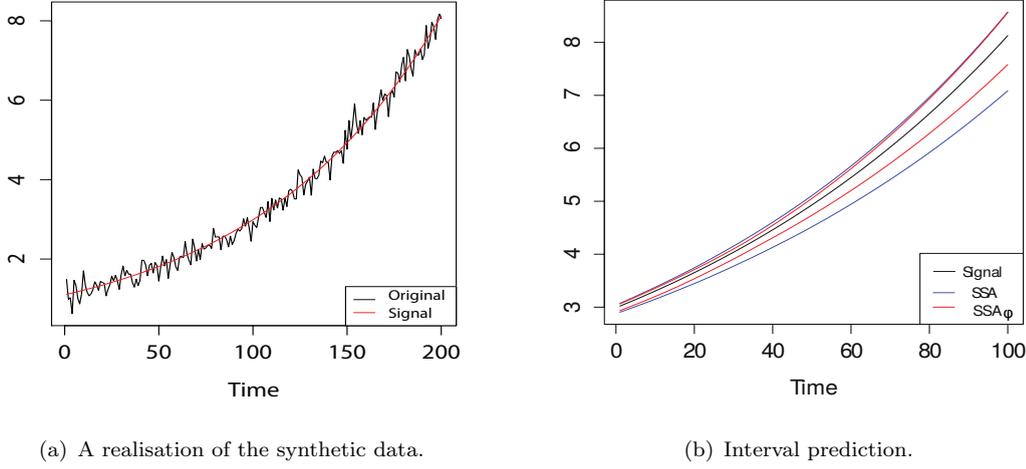


FIGURE 3.1: Realisation of the synthetic data.

Figure 3.1 shows a realisation of the synthetic datasets and compares the FI generated by bootstrap SSA and the bootstrapped LRF coefficients for a SNR= 12. As seen in Figure 3.1(b), the FI provided by bootstrap SSA is wider in comparison with its counterpart, the bootstrapped LRF coefficient. In this case, we can therefore conclude that the bootstrapped SSA FI covers most of the series with high a level of uncertainty and achieves a less accurate forecast (based on the RMSE in Table 3.1) whereas the bootstrap LRF coefficient illustrates tighter bounds with a lower level of uncertainty and higher accuracy (Table 3.1).

As stated before, the following measures are used to evaluate the performance of the competing versions of bootstrap SSA. The (prediction) RMSE and the Ratio of RMSE (RRMSE). The ratio of RMSE is defined as:

$$RRMSE = \left(\sum_{i=1}^N (\hat{y}_{t+h,i} - y_{t+h,i})^2 \right)^{\frac{1}{2}} / \left(\sum_{i=1}^N (\hat{\hat{y}}_{t+h,i} - y_{t+h,i})^2 \right)^{\frac{1}{2}} \quad (3.2)$$

where $\hat{y}_{t+h,i}$ is the h -step ahead forecast obtained using bootstrap LRF coefficients, $\hat{\hat{y}}_{t+h,i}$ is the h -step ahead forecast from bootstrap SSA, and N is the size of the out-of-sample data set.

Table 3.1 presents the results for the average bootstrap FI for the synthetic time series. The first column of Table 3.1 shows a range of values for SNR and the second column is for different window lengths. The results confirm that $SSA_{\bar{\phi}}$ outperforms bootstrap SSA in both aspects. In fact, the quality of the forecast with the Bootstrap LRF coefficients is far superior for a noisy time series. That being the case, updating the LRF coefficients improves the forecast accuracy and results in more a reliable FI than for bootstrap SSA.

TABLE 3.1: Post-sample forecast accuracy measures for synthetic exponential time series.

SNR	L	RMSE		RRMSE	Coverage percentage		ASR	
		SSA	$SSA_{\bar{\Phi}}$	$\frac{SSA_{\bar{\Phi}}}{SSA}$	SSA	$SSA_{\bar{\Phi}}$	SSA	$SSA_{\bar{\Phi}}$
15	6	0.28	0.25	0.90	0.68	0.66	0.05	0.19
	12	0.17	0.14	0.83	0.69	0.70	0.08	0.34
	24	0.08	0.06	0.75	0.75	0.73	0.11	0.47
	36	0.08	0.05	0.63	0.76	0.74	0.12	0.47
	48	0.06	0.04	0.64	0.82	0.78	0.17	0.51
12	6	0.36	0.32	0.89	0.78	0.65	0.85	0.74
	12	0.44	0.40	0.91	0.70	0.58	0.86	0.71
	24	0.24	0.18	0.77	0.71	0.61	0.76	0.66
	36	0.26	0.23	0.87	0.56	0.61	0.87	0.70
	48	0.30	0.26	0.87	0.45	0.61	0.88	0.73
10	6	0.43	0.39	0.89	0.70	0.64	0.93	0.84
	12	0.37	0.33	0.89	0.40	0.54	0.93	0.86
	24	0.40	0.35	0.88	0.10	0.32	0.97	0.91
	36	0.45	0.37	0.82	0.11	0.37	0.96	0.88
	48	0.54	0.33	0.60	0.14	0.48	0.96	0.87
7	6	0.26	0.19	0.73	0.82	0.71	0.80	0.54
	12	0.39	0.33	0.84	0.68	0.60	0.07	0.15
	24	0.28	0.25	0.91	0.37	0.60	0.05	0.23
	36	0.26	0.23	0.90	0.29	0.63	0.05	0.24
	48	0.25	0.21	0.85	0.36	0.67	0.05	0.28
5	6	0.31	0.24	0.75	0.87	0.79	0.07	0.26
	12	0.42	0.36	0.86	0.86	0.78	0.15	0.40
	24	0.89	0.79	0.88	0.85	0.67	0.50	0.61
	36	0.23	0.12	0.50	0.86	0.70	0.35	0.50
	48	0.35	0.10	0.30	0.86	0.75	0.26	0.49
4	6	0.28	0.19	0.68	0.83	0.76	0.15	0.45
	12	0.32	0.26	0.80	0.75	0.73	0.07	0.28
	24	0.35	0.26	0.74	0.78	0.70	0.18	0.45
	36	0.30	0.19	0.62	0.78	0.70	0.19	0.45
	48	0.33	0.23	0.69	0.76	0.73	0.14	0.42

Next, the impact of window lengths and noise levels on coverage percentage is examined. As can be observed from Table 3.1, by increasing noise levels the coverage percentage becomes larger, which confirms the sensitivity of bootstrap SSA to the noise level, whereas these values fall smoothly with window length. However, $SSA_{\bar{\Phi}}$ results in larger coverage values, these values being similar for various level of noise. Thus in this case it is empirically observed that $SSA_{\bar{\Phi}}$ is less sensitive to noise in comparison with bootstrap SSA.

Lastly, Table 3.1 compares the ASR of the SSA bootstrap forecast interval with the bootstrapped LRF coefficients. It shows the bootstrapped LRF coefficients result in a greater asymmetry around the time series. In nearly all cases the bootstrapped SSA coefficients result in ASR values closer to 0.5 demonstrating that this approach is superior in terms of the ASR.

The next section examines the performance of these two models with respect to three real data sets.

3.3.2 Real data sets

In this Section, three sets of data (Figure 3.2) with varying characteristics are used to examine the performance of bootstrapped SSA. The first time series is one of the most commonly used time series known as the death series (Golyandina et al., 2001). The death series represents the monthly accidental deaths in the USA between 1973 and 1978 (Figure 3.2(a)). The data has been used by many researchers and can be found in many time series sources (Chatfield, 2013; Brockwell & Davis, 2002; Davis et al., 2006; Hassani, 2007). As shown in (Hassani & Thomakos, 2010), the best window length and number of eigenvalues for the death series are $L = 24$ and $r = 13$ respectively.

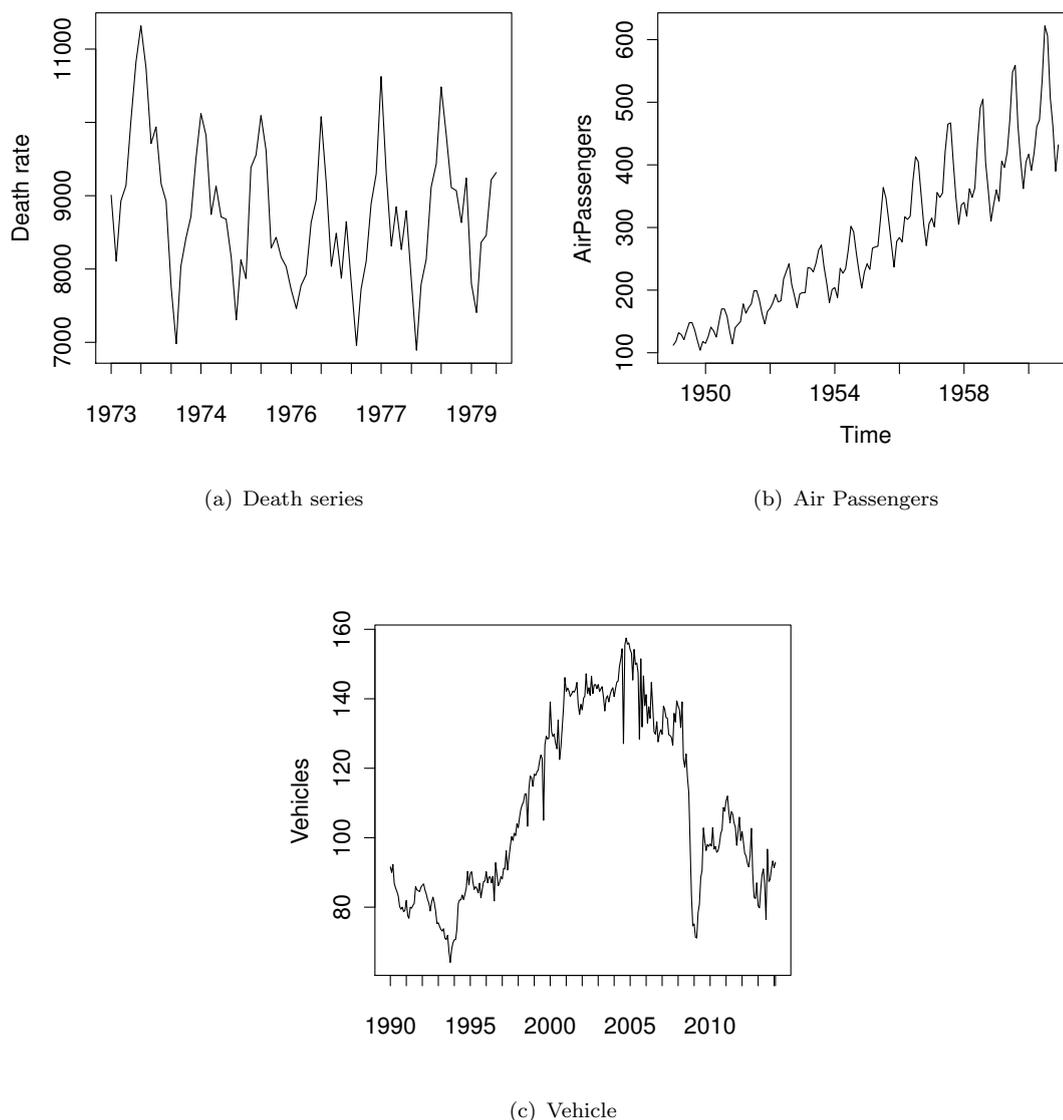


FIGURE 3.2: Three time series used for comparison.

The second time series is the Air Passengers series (Box et al., 2015; Brockwell & Davis, 2002). This series shows monthly totals of international airline passengers between 1949 and 1960 (Figure 3.2(b)). Based on the study by (Mahmoudvand et al., 2013) the optimal parameter values are $L = 22$ and $r = 12$. The third time series is a seasonally adjusted series which measure the production output in the manufacturing of Vehicles in France (Figure 3.2(c)) (Note: Industrial Production series will be explained in detail in Section 4.3.2).

The Death time series exhibits a strong seasonal pattern, with the maximum (peaks) for each year occurring in July and the minimum (troughs) for each year occurring in February (Brockwell & Davis, 2002). The presence of a trend in Figure 3.2(a) is much less apparent than in the Air Passengers series. The Air Passengers series shows both a strong annual pattern and a nearly linear trend. In addition, the variability of the series increases with time, it is suggested by (Brockwell & Davis, 2002) that in this case a logarithmic transformation of the data be used to equalise the variance. The vehicle production series shows a sudden drop after a long period of growth. This *structural break* transfers the underlying system from one state to another which brings some uncertainty about the performance of the model pre-break and after break. These three time series show different combinations of time series components such as a trend, oscillatory components and noise and are thus ideal for comparing the bootstrapped models.

It is typical that forecast errors from differing models are cross-correlated breaking the assumptions inherent in many paired t-tests. Therefore, the Granger-Newbold (GN) test (Mizrach, 1996) which tests not the errors but the differences between forecast errors for significance is employed². The forecasting accuracy is also assessed for four different horizons, 1-step ahead, 3 and 6-steps ahead and one year ahead (12-step). Table 3.2 summarises the post-sample forecast accuracy measured by the two competing models over the three data sets.

In Table 3.2 the bootstrapped LRF coefficient model performs better (wrt RMSE) in all cases with 8 out of 12 figures being statistically significant ($\alpha = 0.01$). It is interesting to note that the results are not significant for the French vehicle production data which may be due to the structural break as will next be examined. Finally, we note that although the results are significant they might not be considered important in these cases.

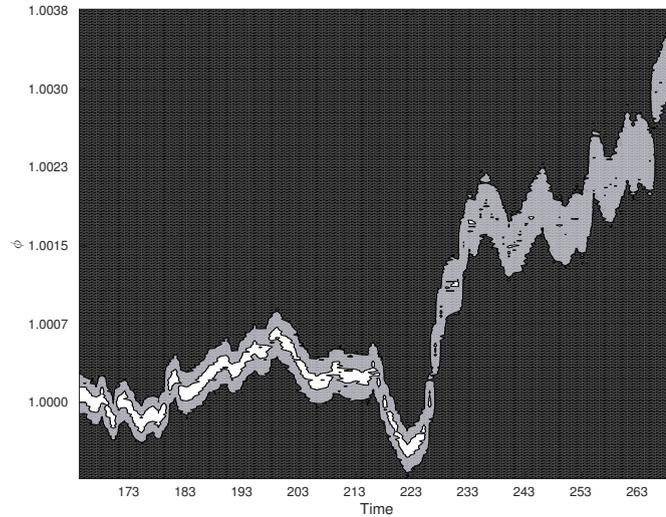
As mentioned before, there is an unexpected shift in the vehicle time series which can lead to uncertainty and unreliability with SSA based forecasting. Therefore, more investigation is needed to evaluate the performance of the core of the SSA forecasts, specifically the coefficients of the LRF, in the presence of a structural break. Figure 3.3 illustrates

²More details can be found in Appendix A.

TABLE 3.2: Post-sample forecast accuracy measures for the real time series.

Series	h	RMSE			RRMSE	
		SSA	Boot SSA	$SSA_{\bar{\phi}}$	$\frac{BootSSA}{SSA}$	$\frac{SSA_{\bar{\phi}}}{SSA}$
Death	1	521.00	520.10	502.20	1.00	0.97
	3	604.31	599.37	568.10	0.99	0.95 *
	6	629.20	604.44	570.50	0.96 *	0.94 *
	12	775.59	736.50	560.20	0.95 *	0.76 *
Air Passengers	1	22.36	22.36	20.96	1.00	0.94 *
	3	24.28	24.29	23.33	1.00	0.96 *
	6	28.98	28.93	27.94	1.00	0.97 *
	12	40.25	40.20	38.25	1.00	0.95 *
Vehicle	1	6.50	6.50	6.20	1.00	0.95 *
	3	10.34	10.35	10.29	1.00	0.99
	6	15.50	15.53	15.43	1.00	0.99
	12	21.69	21.74	21.15	1.00	0.97

Note: * indicates results are statistically significant at $\alpha=0.01$ based on modified GN test.

FIGURE 3.3: Cloud map of ϕ_1 for the Vehicle time series.

a cloud map of the LRF coefficients.³ It is evident that, before the structural break (at time 226), the average is changing smoothly and the distribution of coefficients is more concentrated around the mean (white area). After the break, the distribution becomes wider and less concentrated around the mean (the colours are less intense and the band is wider). Subsequently, this greater uncertainty in ϕ will result in greater uncertainty in the forecasts. What interests us most is that this uncertainty appears to persist as times move on. It is therefore important to investigate the underlying causes of this discrepancy (Chapter 4).

³This figure shows 100 histograms with the same intervals in each, and the frequency of each bin is coloured appropriately to aid visualisation of how these change over time.

3.4 Conclusion

The idea of SSA forecasting utilising the bootstrapped LRF coefficients appears empirically to give some promising results which warrants further research. Bootstrap SSA as a benchmark model, was considered to compare the performance of the proposed method. The sensitivity of the technique, with respect to different noise levels and window lengths, was evaluated by using both synthetic and real time series.

However, from the viewpoint of this research, the main finding of this chapter is that, when a shock happens to a system, the coefficients of the LRF may become less concentrated and volatile. This phenomenon is established by observing that these changes tend to be persistent over time and therefore feed more uncertainty into the forecasts. The key question is whether this is due to the true underlying distribution of the states in the system being more volatile or whether the model has failed to account for a structural break correctly. Finally, in the next chapter, we look specifically at SSA in the presence of a structural break.

Chapter 4

Bayesian Singular Spectrum Analysis with a State Dependent model

4.1 Introduction

As noted by (Golyandina et al., 2001; Rahmani et al., 2016), in certain circumstances, such as in the presence of a structural break, turning points or complex cyclical patterns, an LRF's parameters may not match a linear continuation of the time series pre-break. Further in Chapter 3 we saw that the distribution of the coefficients of the SSA algorithm become less concentrated around their mean and that this effect is persistent. However, with SSA a time series is assumed to be one realisation from the same LRF; i.e. the fact that the series is historically different prior to a structural break is not taken into account. However, even though the LRF prior to the break is different from that post break there still remains commonalities between the two; i.e. we assume that in some fashion the time series follows the same rules (for example, the economy after a crash still evolves according to economics). SSA assumes that a time series evolves according to an underlying state space system. A structural break is equivalent to a persistent shock to that system (for example a crash in economic activity persists over time) that moves those states from one region of the state space to another region over a short period of time; thereafter the system then continues to evolve (by assumption) according to the (unknown) state transition function as before. Although the system post-break evolves according to the same dynamics, basic SSA is unable to estimate these parameters correctly (as we shall prove in Theorem 4.1) and in essence we require not a fixed state space system as in SSA but one which adapts to the jump in the region of the state space. Such a model is known as a *state-dependent model*.

For this to happen, the movement of LRF parameters are tracked by using a state-dependent model (Priestley, 1980) in such a way that the parameters can change not only over time but also over the states of the system. To put it differently, a non-linear stochastic dynamic system is explained by a non-linear function connecting the current states of a variable to its previous states (and possibly the previous states of other series) and random shocks. The whole process is called Bayesian MSSA since the parameters are no longer fixed for all time series and will be updated recursively via EKF.

It will be shown that this new framework of parameters provides MSSA with more flexibility to cover non-stationarity, non-linearity and structural breaks in time series, in which the transition matrix essentially relies on the inferred evolution of the states over time.

The proposed model is tested on synthetic data (Univariate and Multivariate) and the monthly Industrial Production Indices (IPI) of France, Germany and the UK. To provide a better understanding of its performance regarding both univariate SSA and classical autoregressive models, the results are compared with ARIMA, ARFIMA, ETS, GARCH, VAR, VECM, SSA and MSSA.

4.2 The Multivariate Bayesian SSA algorithm

4.2.1 SSA model

SSA is a non-parametric technique based on the covariance structure of a time series. The covariance structure is captured using an embedded representation of a time series which decomposes it into a set of data-adaptive orthonormal components. These components can then be projected into a lower dimension via SVD which produces a smoother time series which can be used for forecasting. Below we make specific the various elements from Section 2.1 with respect to the Bayesian MSSA algorithm.

Let $\mathbf{Y} = \{Y_N^{(m)} = (y_1^{(m)}, \dots, y_N^{(m)}) : m = 1, \dots, M\}$ be a multivariate time series with M series (channels) of length N . In the first step, windows of length L of $Y_N^{(m)}$, $\mathbf{X}_i^{(m)} = [y_i^{(m)} \dots y_{i+L}^{(m)}]^T$, $i = 1, \dots, N - L + 1$ are used to embed the series as $\mathbf{X}^{(m)} = [\mathbf{X}_1^{(m)}, \dots, \mathbf{X}_{N-L+1}^{(m)}] \in \mathcal{R}^{N-L+1 \times L}$ where $\mathbf{X}^{(m)}$ is called the *block trajectory* matrix of $Y_N^{(m)}$ with accompanying *trajectory space* $\mathcal{L}_L = \text{span}\{\mathbf{X}_1, \dots, \mathbf{X}_{N-L+1}\}$. In univariate SSA the SVD of $\mathbf{X}^{(m)}$ is used to determine the eigenstructure of this block of delayed versions of the time series, i.e. its covariance structure. In MSSA the common covariance structure of multiple time series is uncovered through their stacked Hankel matrix,

$\mathbf{X}_M = [\mathbf{X}^{(1)} \mathbf{X}^{(2)} \dots \mathbf{X}^{(M)}]$, as:

$$\mathbf{X}_M = \begin{pmatrix} y_1^{(1)} & y_2^{(1)} & \cdots & y_{N-L+1}^{(1)} & \cdots & y_1^{(M)} & y_2^{(M)} & \cdots & y_{N-L+1}^{(M)} \\ y_2^{(1)} & y_3^{(1)} & \cdots & y_{N-L+2}^{(1)} & \cdots & y_2^{(M)} & y_3^{(M)} & \cdots & y_{N-L+2}^{(M)} \\ \vdots & \vdots \\ y_{L-1}^{(1)} & y_L^{(1)} & \cdots & y_{N-1}^{(1)} & \cdots & y_{L-1}^{(M)} & y_L^{(M)} & \cdots & y_{N-1}^{(M)} \\ y_L^{(1)} & y_{L+1}^{(1)} & \cdots & y_N^{(1)} & \cdots & y_L^{(M)} & y_{L+1}^{(M)} & \cdots & y_N^{(M)} \end{pmatrix}$$

The next step involves seeking an embedding of \mathbf{X}_M which according to Taken's theorem exists if the time series may be described by a finite set of embedded state vectors; further the SVD may be used to approximate such an embedding (Golyandina et al., 2001).

Given $C_X = \mathbf{X}_M \mathbf{X}_M^T \in \mathcal{R}^{L \times L}$, the covariance matrix of the rows of \mathbf{X}_M , then denote $U_{M_j} = (u_{j1}, \dots, u_{Lj})^T$ as the left singular vectors of \mathbf{X}_M with λ_j as the corresponding eigenvalues. An embedding is produced by selecting $r < L$ eigen components, where $\hat{\mathbf{X}}_M = [\hat{\mathbf{X}}^{(1)} \hat{\mathbf{X}}^{(2)} \dots \hat{\mathbf{X}}^{(M)}] = \sum_{j=1}^r U_{M_j} U_{M_j}^T \mathbf{X}_M$ describes the reconstructed matrix. The SVD is optimal in the sense that among all the matrices of rank r , the matrix $\hat{\mathbf{X}}_M$ provides the best approximation to the trajectory matrix \mathbf{X}_M in the norm sense, such that $\|\mathbf{X}_M - \hat{\mathbf{X}}_M\|_{\mathcal{M}}$ is a minimum, where \mathcal{M} is the collection of matrices, see (Golyandina et al., 2001; Patterson et al., 2011). To form the reconstructed times series, a Hankelization process (diagonal averaging) can be applied on each of the blocks, $\tilde{\mathbf{X}}^{(m)} = \mathcal{H}\hat{\mathbf{X}}^{(m)}$; ($m = 1, \dots, M$), which transforms them into a new time series of the same length by the following formula:

$$\tilde{y}_k^{(m)} = \begin{cases} \frac{1}{k} \sum_{l=1}^k \hat{x}_{l, k-l+1}^{(m)}, & 1 \leq k < L^*, \\ \frac{1}{L^*} \sum_{l=1}^{L^*} \hat{x}_{l, k-l+1}^{(m)}, & L^* \leq k \leq K^*, \\ \frac{1}{N-k+1} \sum_{l=k-K^*+1}^{N-K^*+1} \hat{x}_{l, k-l+1}^{(m)}, & K^* < k \leq N. \end{cases} \quad (4.1)$$

where $K^* = \max(L, K)$ and $L^* = \min(L, K)$.

To produce a forecast, an underlying assumption is that $\tilde{y}_k^{(m)}$ satisfies an LRF. Assume $\mathcal{L}_r = \text{span}\{U_{M_1}, \dots, U_{M_r}\}$ then by grouping of the elementary matrices corresponding to the set of $I = \{J_1, \dots, J_r\} \in \{1, \dots, L\}$, the series can be governed by the LRF as:

$$\begin{aligned} \hat{y}_k^{(m)} &= \phi_1 \hat{y}_{k-1}^{(m)} + \dots + \phi_{L-1} \hat{y}_{k-L+1}^{(m)} \\ \Leftrightarrow \mathbf{X}^{(m)} &= \left(\begin{array}{c|ccc} \mathbf{0} & & & \mathbf{I} \\ \hline 0 & \phi_1 & \cdots & \phi_{L-1} \end{array} \right) \mathbf{X}^{(m-1)} \end{aligned} \quad (4.2)$$

which is also defined as an L -continuation of the series where the coefficients vector are applied to the linear space \mathcal{L}_r . Let $U_{M_j}^\nabla$ denote the vector of the first $L-1$ coordinates of the eigenvectors U_{M_j} , and π_{M_j} denotes the last coordinate of the eigenvectors U_{M_j} , then

if $e_L \notin \mathcal{L}_r$ it can be simply proved that $\pi_{M_1}^2 + \pi_{M_2}^2 + \dots + \pi_{M_r}^2 < 1$, see (Golyandina et al., 2001). It is also shown that there exists a unique vector $(y_{k-1}^{(m)}, \dots, y_{k-L+1}^{(m)}, y_\tau^{(m)}) \in \mathcal{L}_r$ which corresponds to any vector $(y_{k-1}^{(m)}, \dots, y_{k-L+1}^{(m)}) \in \mathcal{L}_r^\nabla$ with numbers h such that

$$(y_{k-1}^{(m)}, \dots, y_{k-L+1}^{(m)}, y_\tau^{(m)})^T = (y_{k-1}^{(m)}, \dots, y_{k-L+1}^{(m)}, 0)^T + y_\tau^{(m)} e_L = \sum_{j=1}^r h_j U_{M_j}.$$

Multiplying both sides by e_L and U_{M_j} respectively, results in

$$y_\tau^{(m)} = \sum_{j=1}^r h_j \pi_j, \quad \text{and} \quad \sum_{j=1}^r h_j = \sum_{j=1}^r (y_{k-1}^{(m)}, \dots, y_{k-L+1}^{(m)})^T U_{M_j}^\nabla + y_\tau^{(m)} \sum_{j=1}^r \pi_j.$$

By substitution:

$$\begin{aligned} y_\tau^{(m)} &= \sum_{j=1}^r h_j \pi_j = \sum_{j=1}^r (y_{k-1}^{(m)}, \dots, y_{k-L+1}^{(m)})^T \pi_j U_{M_j}^\nabla + y_\tau^{(m)} \sum_{j=1}^r \pi_j^2 \\ &= (1 - \sum_{j=1}^r \pi_j^2)^{-1} \left[\sum_{j=1}^r (y_{k-1}^{(m)}, \dots, y_{k-L+1}^{(m)})^T U_{M_j}^\nabla \pi_j \right] \end{aligned}$$

Therefore, the linear parameters (or coefficients), $\Phi = (\phi_{L-1}, \dots, \phi_1)^T$, of equation (4.2) can be expressed as:

$$\Phi = \frac{1}{1 - \sum_{j=1}^r \pi_j^2} \sum_{j=1}^r \pi_{M_j} U_{M_j}^\nabla \quad (4.3)$$

and forecasts at time τ ($\tau \geq k$) can be straightforwardly obtained by substitution of the reconstructed time series when known or alternatively the forecasts, i.e.:

$$[\hat{y}_\tau^{(1)}, \dots, \hat{y}_\tau^{(M)}]^T = \begin{cases} [\tilde{y}_\tau^{(1)}, \dots, \tilde{y}_\tau^{(M)}]^T, & \tau = 1, \dots, k-1, \\ \Phi^T \begin{bmatrix} (\hat{y}_{\tau-L+1}^{(1)}, \dots, \hat{y}_{\tau-1}^{(1)}) \\ \vdots \\ (\hat{y}_{\tau-L+1}^{(M)}, \dots, \hat{y}_{\tau-1}^{(M)}) \end{bmatrix}, & \tau = k, \dots, N. \end{cases} \quad (4.4)$$

4.2.2 Linear Recurrent Formula in the presence of a structural break

In the presence of a structural break, estimates for Φ can substantially change as the series changes, and as the following theorem shows the LRF post a shift in the mean does not correspond to linear continuation of the time series pre-shift.

Theorem 4.1. *LRF Theorem* Let $\check{Y} \in \mathcal{R}^{N \times M}$ be a multivariate time series which follows an LRF. Assume that the observed time series is a back-shifted version of $\check{Y} \in \mathcal{R}^{N \times M}$ such that $Y_m = \check{Y}_m - \delta_m \forall t < \tau_m$, where each time series experiences a structural shift in the mean, δ_m , at time τ_m . Let C_X and \check{C}_X denote the covariance matrices of the

trajectories of Y and \check{Y} respectively. Then the deviation between the first eigenvalue of C_X and \check{C}_X is $\lambda_1 - \check{\lambda}_1 \approx \mathcal{T}L(\mathcal{Q}^2 - 2\mathcal{Q}\bar{y})$. Furthermore, this deviation decreases as $\mathcal{O}(k/N)$.

Proof. First note that for $t > \max(\tau_m)$ all structural shifts have occurred and the observed time series (by assumption) is a linear continuation of \check{Y} , not Y . Thus \check{C}_X contains the correct covariance structure for the time series going forward.

The covariance matrix of the trajectory matrix can be explicitly written in terms of the original time series without any structural shifts, $y_k^{(m)}$, as:

$$\check{C}_X = \begin{pmatrix} \sum_{j=1}^M \sum_{i=1}^{N-L+1} y_i^{(m)2} & \sum_{j=1}^M \sum_{i=1}^{N-L+1} y_i^{(m)} y_{i+1}^{(m)} & \cdots & \sum_{j=1}^M \sum_{i=1}^{N-L+1} y_i^{(m)} y_{i+L-1}^{(m)} \\ \sum_{j=1}^M \sum_{i=1}^{N-L+1} y_i^{(m)} y_{i+1}^{(m)} & \sum_{j=1}^M \sum_{i=2}^{N-L+2} y_i^{(m)2} & \cdots & \sum_{j=1}^M \sum_{i=2}^{N-L+2} y_i^{(m)} y_{i+L-2}^{(m)} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^M \sum_{i=L-1}^{N-1} y_i^{(m)} y_{i-(L-2)}^{(m)} & \cdots & \sum_{j=1}^M \sum_{i=L-1}^{N-1} y_i^{(m)2} & \sum_{j=1}^M \sum_{i=L-1}^{N-1} y_i^{(m)} y_{i+1}^{(m)} \\ \sum_{j=1}^M \sum_{i=L}^N y_i^{(m)} y_{i-(L-1)}^{(m)} & \cdots & \sum_{j=1}^M \sum_{i=L}^N y_i^{(m)} y_{i-1}^{(m)} & \sum_{j=1}^M \sum_{i=L}^N y_i^{(m)2} \end{pmatrix}$$

Before proceeding we require the following lemma.

Lemma 4.2 (Lemma). *If the time series, $y_i^{(m)}$, contains a constant term, $c^{(m)}$, then \check{C}_X has an eigenvector $u_1 \approx [1, 1, \dots, 1]^T 1/\sqrt{L}$ with corresponding eigenvalue $\lambda_1 \approx L(N-L+1) \sum_m c^{(m)2}$.*

Proof. Each entry in \check{C}_X consists of a sum of $(N-L+1)$ products of the time series with itself. In the presence of a constant, $E[y_i^{(m)} y_{i-\tau}^{(m)}] \approx c^{(m)2}$ for some delay τ . Thus, $\check{C}_X \approx (N-L+1) \sum_m c^{(m)2} \mathbf{J}_{L \times L}$, where $\mathbf{J}_{L \times L}$ is a matrix of ones. $\mathbf{J}_{L \times L}$ has only one eigenvector $[1, 1, \dots, 1]^T 1/\sqrt{L}$ and this completes the proof. \square

Returning to the observed time series, note that C_X similarly consists of two types of terms, sums of squares along the diagonal and sums of cross-terms along the off-diagonal elements. There are, in addition, shifts in the mean and these terms may therefore be expressed as:

$$c_{i,j} = \begin{cases} \sum_{m=1}^M \sum_{k=l}^{N-L+l} (y_k^{(m)} - \delta_m \mathbf{1}_{A_m})^2, & i = j, \\ \sum_{m=1}^M \sum_{k=l}^{N-L+l} (y_k^{(m)} - \delta_m \mathbf{1}_{A_m})(y_{k+|i-j|}^{(m)} - \delta_m \mathbf{1}_{A_m}), & i \neq j. \end{cases} \quad (4.5)$$

where $l = \min(i, j)$ and $\mathbf{1}_{A_m}$ is a step function with $\mathbf{1}_{A_m} = 1$ for $A_m = \{t < \tau_m\}$ and zero otherwise. Expanding these terms results in:

$$c_{i,j} = \begin{cases} \sum_{m=1}^M \tau_m \delta_m^2 - 2 \sum_{m=1}^M \tau_m \delta_m \sum_{k=l}^{N-L+l} y_k^{(m)} + \sum_{m=1}^M \sum_{k=l}^{N-L+l} y_k^{(m)2}, & i = j, \\ \sum_{m=1}^M \tau_m \delta_m^2 - \sum_{m=1}^M \tau_m \delta_m \left(\sum_{k=l}^{N-L+l} y_k^{(m)} + \sum_{k=l}^{N-L+l} y_{k+|i-j|}^{(m)} \right) + \sum_{m=1}^M \sum_{k=l}^{N-L+l} y_k^{(m)} y_{k+|i-j|}^{(m)}, & i \neq j. \end{cases} \quad (4.6)$$

Now assume the time series is not changing rapidly or that there is a constant term which dominates (and thus is separable) then the expected sum of the time series in overlapping windows is approximately constant as:

$$\sum_{m=1}^M \tau_m \delta_m \left(\sum_{k=l}^{N-L+l} y_k^{(m)} + \sum_{k=l}^{N-L+l} y_{k+|i-j|}^{(m)} \right) \simeq 2 \sum_{m=1}^M \tau_m \delta_m \sum_{k=l}^{N-L+l} y_k^{(m)} \quad (4.7)$$

and so it can be seen that \check{C}_X is related to C_X as:

$$c_{i,j} \approx \check{c}_{i,j} + \sum_{m=1}^M \tau_m \delta_m^2 - 2 \sum_{m=1}^M \tau_m \delta_m \sum_{k=l}^{N-L+l} y_k^{(m)} \quad (4.8)$$

then the covariance matrix can be rewritten as:

$$C_X \simeq \check{C}_X + \left(\mathcal{T} \mathcal{Q}^2 - 2 \mathcal{T} \mathcal{Q} \sum_{k=l}^{N-L+l} y_k \right) \mathbf{J}_{L \times L} \approx \check{C}_X + \mathcal{T} L (\mathcal{Q}^2 - 2 \mathcal{Q} \bar{y}) \frac{1}{\sqrt{L}} \mathbf{1}_L \mathbf{1}_L^T \frac{1}{\sqrt{L}} \quad (4.9)$$

where \mathcal{T} is a row vector with the numbers of samples prior to the structural breaks, \mathcal{Q} is a column vector of amplitudes of the structural breaks, $\mathbf{1}_L$ is vector of ones with L elements, and $\mathbf{J}_{L \times L}$ is a square matrix of ones. Thus the net effect of a structural break is that the first eigenvalue of the covariance matrix has a bias of approximately $\mathcal{T} L (\mathcal{Q}^2 - 2 \mathcal{Q} \bar{y})$. From Lemma 4.2.2 this bias can be seen to tend to zero with N as:

$$\frac{\lambda_1 - \check{\lambda}_1}{\check{\lambda}_1} \approx \frac{\mathcal{T} L (\mathcal{Q}^2 - 2 \mathcal{Q} \bar{y})}{(N - L + 1) \sum_m c^{(m)2}} = \mathcal{O}(k/N) \quad (4.10)$$

where k is some constant. □

Theorem 4.1 states that the first eigenvalue of C_X will contain a bias which slowly decays as new samples arrive (See Figure 4.2(b) for an empirical example of how this effects the LRF parameter estimates). For this reason a dynamic LRF is proposed based on a state dependent model.

4.2.3 State dependent format of Linear Recurrent Formula parameters

In this Section, the LRF parameters are viewed as state parameters which we allow to recursively evolve based on the observations. Define $Y_{k-1}^{(m)} = \{y_{k-L+1}^{(m)}, \dots, y_{k-1}^{(m)}\}^T$ as the state vector of the m^{th} time series. In a state dependent model (Priestley, 1980) the propagation coefficients are a function of the state vector as:

$$y_k^{(m)} = \phi_1(Y_{k-1}^{(m)})y_{k-1}^{(m)} + \phi_2(Y_{k-1}^{(m)})y_{k-2}^{(m)} + \dots + \phi_{L-1}(Y_{k-1}^{(m)})y_{k-L+1}^{(m)} \quad (4.11)$$

Note that in contrast with MSSA, parameters estimated by MBSSA (above) have different values for each series. In addition, $\{\phi_u(Y_{k-1}^{(m)})\} = \{\phi_u^{(m)}\}$, are assumed to be analytic functions of $y_k^{(m)}$ which change smoothly over time and so can be expressed via a Taylors series expansion as:

$$\phi_{u,k}^{(m)} = \phi_{u,k-1}^{(m)} + \Delta y_{k-u}^{(m)} \gamma_{k,u}^{(m)}, \quad u = 1, \dots, L-1, \quad (4.12)$$

where $\Delta y_{k-u}^{(m)} = y_{k-u}^{(m)} - y_{k-(u-d)}^{(m)}$, $\gamma_{k,u}$ is a gradient, and d is the seasonal length of the time series. The gradient parameters $\gamma_{k,1}^{(m)}, \dots, \gamma_{k,L-1}^{(m)}$ are unknown however, and, as they are hyperparameters it can be assumed that they change slowly following a random walk as:

$$\Gamma_{k+1}^{(m)} = \Gamma_k^{(m)} + V_{k+1}^{(m)}, \quad (4.13)$$

while $\Gamma_k^{(m)} = (\gamma_{k,1}^{(m)}, \dots, \gamma_{k,L-1}^{(m)})$ and $V_{k+1}^{(m)}$ is a sequence of independent matrix-valued random variables such that $V_{k+1}^{(m)} \sim N(0, \Sigma_{V_{k+1}})$. With some modifications, the general recursive model can be rewritten in a state-space form in which the state-vector is no longer $Y_k^{(m)}$, but is replaced by the state-dependent coefficients augmented with the gradients:

$$\theta_k^{(m)} = [\phi_{k-1,1}^{(m)}, \dots, \phi_{k-1,L-1}^{(m)}, \gamma_{k,1}^{(m)}, \dots, \gamma_{k,L-1}^{(m)}]^T, \quad (4.14)$$

where $\theta_k^{(m)}$ is the vector of all unknown parameters of the model (and of course a function of $Y_{k-1}^{(m)}$). To be more specific, the SDM scheme can be reformulated in a new state space format by replacing the states with equation (4.14) as:

$$\begin{aligned} y_k &= H_k^* \theta_k^{(m)} + \epsilon_k, \\ \theta_k^{(m)} &= \mathbf{F}_{k-1}^* \theta_{k-1}^{(m)} + W_k. \end{aligned} \quad (4.15)$$

where H_k^* is an inherent characteristic of the system capturing its motion and can be considered as $H_k^* = (y_{k-1}, \dots, y_{k-L+1}, 0, \dots, 0)$. $\mathbf{F}_k^* = \mathbf{F}_k^*(Y_k^{(m)})$ is a state transition matrix and is naturally characterised into a corresponding block diagonal form and open to modification over time. W_k is evolution noise or innovation and uncorrelated to θ_k . Technically, there is no need to assume a Gaussian distribution governs ϵ_t , however, defining a prior distribution is necessary to produce initial values for the state vector (Steel, 2001).

Equation (4.15) presents the standard form to which the Extended Kalman Filter (EKF) can be applied directly as:

$$\hat{\theta}_k^{(m)} = \mathbf{F}_{k-1}^* \theta_{k-1}^{(m)} + \mathbf{K}_k^* (y_k^{(m)} - H_k^* \mathbf{F}_{k-1}^* \theta_{k-1}^{(m)}), \quad (4.16)$$

where the $2(L-1) \times 2(L-1)$ matrix \mathbf{F}_{k-1}^* is given by:

$$\mathbf{F}_{k-1}^* = \left(\begin{array}{c|ccc} & \Delta y_{k-1} & & 0 \\ & & \Delta y_{k-2} & \\ \mathbf{I}_{L-1} & & & \ddots \\ & & & & \Delta y_{k-(L-1)} \\ \hline 0 & & & \mathbf{I}_{L-1} & \end{array} \right)$$

and \mathbf{K}_k^* is the Kalman gain matrix where $\mathbf{K}_k^* = \mathbf{\Phi}_k (H_k^*)^T \sigma_e^2$, and $\mathbf{\Phi}_k$ is the variance-covariance matrix of the one-step prediction error of θ_k , *i.e.* $\mathbf{\Phi}_k = E[(\theta_k - \mathbf{F}_{k-1}^* \hat{\theta}_{k-1})(\theta_k - \mathbf{F}_{k-1}^* \hat{\theta}_{k-1})^T]$. Moreover, σ_e^2 is the variance of the one-step ahead prediction error of y_k , *i.e.*, σ_e^2 is the variance of $e_k = y_k - H_k^* \mathbf{F}_{k-1}^* \hat{\theta}_{k-1}$. Denote \mathbf{C}_k as the variance-covariance matrix of $(\theta_k - \hat{\theta}_k)$, then successive values of $\hat{\theta}_k$ may be estimated by using the standard recursive algorithm of the Kalman filter:

$$\begin{aligned} \mathbf{\Phi}_k &= \mathbf{F}_{k-1}^* \mathbf{C}_{k-1} (\mathbf{F}_{k-1}^*)^T + \Sigma_W, & \Sigma_W &= \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_V \end{pmatrix}, \\ \mathbf{K}_k^* &= \mathbf{\Phi}_k (H_k^*)^T \left[H_k^* \mathbf{\Phi}_k (H_k^*)^T + \sigma_e^2 \right]^{-1}, \\ \mathbf{C}_k &= \mathbf{\Phi}_k - \mathbf{K}_k^* \left[H_k^* \mathbf{\Phi}_k (H_k^*)^T + \sigma_e^2 \right] (\mathbf{K}_k^*)^T, \end{aligned} \quad (4.17)$$

To implement the foregoing algorithm, initial values need to be provided. Many different ways of doing this have been suggested, for instance (Priestley, 1980) recommends taking an initial stretch of the data and fitting a standard AR model. In this study, bootstrapped SSA (Golyandina et al., 2001) is used to generate initial estimates of θ_{k_0-1} , σ_e^2 and C_{k_0-1} . MSSA is initially used to extract the signal, the residual and thus estimate the distribution of the noise. Bootstrapped samples are then drawn from this distribution, added to the signal and this gives bootstrapped estimates for θ_k , σ_e^2 and C_k .¹

¹Note that it is assumed the initial gradients to be zero, and in addition to be independent of ϕ_k .

As soon as the Kalman filtering recursion starts, the effect of the chosen starting value is seemingly negligible with the initial estimated coefficients as: $\hat{\theta}_{k_0-1} = (\hat{\phi}_{k_0,1}, \dots, \hat{\phi}_{k_0,L-1}, 0, \dots, 0)$ and the residual variance of the model $\hat{\sigma}_\epsilon^2$ and $\mathbf{C}_{k_0-1} = \begin{pmatrix} \hat{\mathbf{\Omega}}_{\phi_1, \dots, \phi_{L-1}} & 0 \\ 0 & 0 \end{pmatrix}$, where $\hat{\mathbf{\Omega}}$ is the estimated variance-covariance matrix of $(\hat{\phi}_{k_0,1}, \dots, \hat{\phi}_{k_0,L-1})$ gained from bootstrapping. The selected range for the smoothing factor in this study is between 10^{-3} to 10^{-6} .

4.3 Empirical results; Univariate

To motivate the contribution of the Bayesian LRF to forecasting time series with shifts and structural breaks, a number of experiments are performed using both synthetic and real data. Both univariate SSA and BSSA are tested on synthetic data and the results are discussed in the next section. The second part presents the forecasting results for real data (Industrial Production Indicators) at horizons of up to a year.

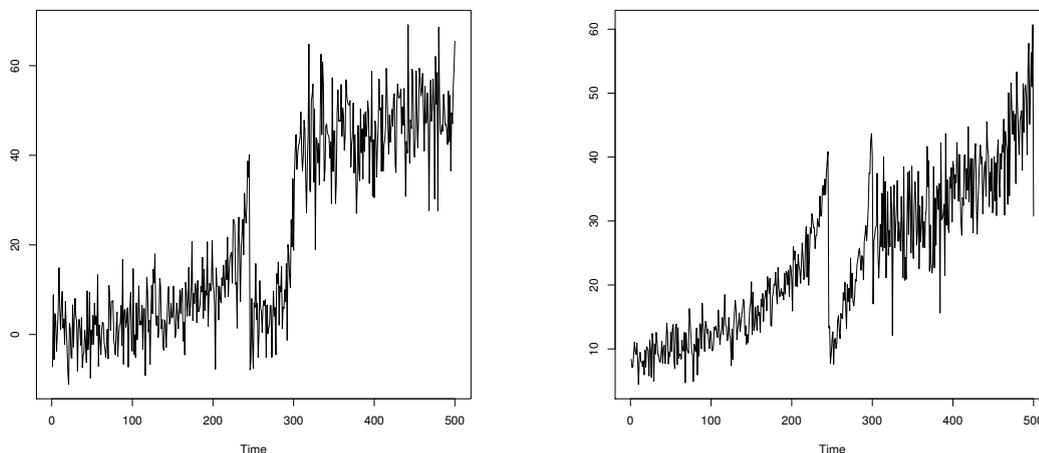
4.3.1 Synthetic data

The synthetic signals are formed as a mixture of two signals $Y_N(t) = S_N(t) + \beta E_N(t)$, where S_N indicates the signal to be extracted, E_N is white noise and β is the noise level. We tested the following synthetic model which is a general, non-linear model and consists of an exponential component to model growth, a sinusoidal component to model seasonality and mean component through which structural change will be introduced. Specifically, the signal $S_N(t)$ has the following form:

$$S_N(t) = \theta e^{-\nu t} + \eta \sin(\omega t) + \mu \quad (4.18)$$

where $\theta \in \mathbb{R}$ is a constant, $\nu \in \mathbb{R}$ is growth rate, $\eta \in \mathbb{R}$ is amplitude, $\omega \in (0, 2\pi)$ is the angular frequency, and μ is vertical shift or the mean level of the signal. Note that different values of $(\beta, \theta, \nu, \eta, \omega, \mu)$ generate time series with different characteristics and behaviour.

We conducted two sets of simulation experiments to evaluate the performance of the models. In the first set of simulations (Figure 4.1(a)), structural breaks in the mean are introduced at times 250 and 300. While in the second set of simulations (Figure 4.1(b)), the true process contains structural breaks which change the variance of the series with time in order to add instability to the data. Figure 4.1 shows two realisations of the synthetic data sets in which the turning point changes mean and variance level of the data at time point $Q = 250$. We then compare the performance of SSA and BSSA following both structural breaks, i.e. from time 250 to 500, over 200 realisations of each process.



(a) Shift in mean level: $N = 500, \theta = 0.5, \eta = 2, \nu = 0.04, \omega = \pi/12, \beta = 5$ and μ equals to 0, 10 and 30 for each stages. (b) Shift in variance: $N = 500, \theta = 0.5, \eta = 2, \nu = 0.01, \omega = \pi/12, \mu = 10$ and β equals to 4, 8 and 15.

FIGURE 4.1: Realisation of synthetic data.

As can be seen from Figure 4.1, these jumps move the time series from one state to another state such that the LRF governing the forecast series can be affected and displays different behaviour afterwards. Figure 4.2 shows the leading coefficients of the LRF estimated by SSA and BSSA. As expected, the basic SSA coefficients exhibit a sudden rise from 1 to 1.02 and thereafter follow an exponential decay before returning back to its initial values. In contrast, the BSSA estimates vary considerably along the changes of the series (The same results are also obtained from the real datasets). It is also worth pointing out that the presence of a constant term (vertical shift), μ , in equation (4.18) corresponds to an eigenvector of $[1, 1, \dots, 1]^T 1/\sqrt{L}$ in U_i (see example 3.1 in (Usevich, 2010)). However, the corresponding eigenvalue performs weaker by a factor of Q/N which is due to the absence of μ prior to the structural break. That means if $N \rightarrow \infty$ then $Q/N \rightarrow 0$, and this explains the decay in the empirical coefficient observed in Figure 4.2(b).

To compare the performance of SSA and BSSA precisely, signals (Equation (4.18)) with differing noise levels are generated. A range of β is considered between 4–40 which measures the level of nonstationary Gaussian noise to the signal. In this way, the effect of nonlinearity and nonstationarity when SSA and BSSA are used can be compared. This is useful since many real signals in nature are nonstationary. Figure 4.3 demonstrates the performance of the two models, SSA and BSSA, versus the changes in noise level. Compared to SSA, BSSA performs quite well for both sets of synthetic data. Therefore, it can be concluded that the LRF coefficients need to be adjusted properly in the presence of a structural break.

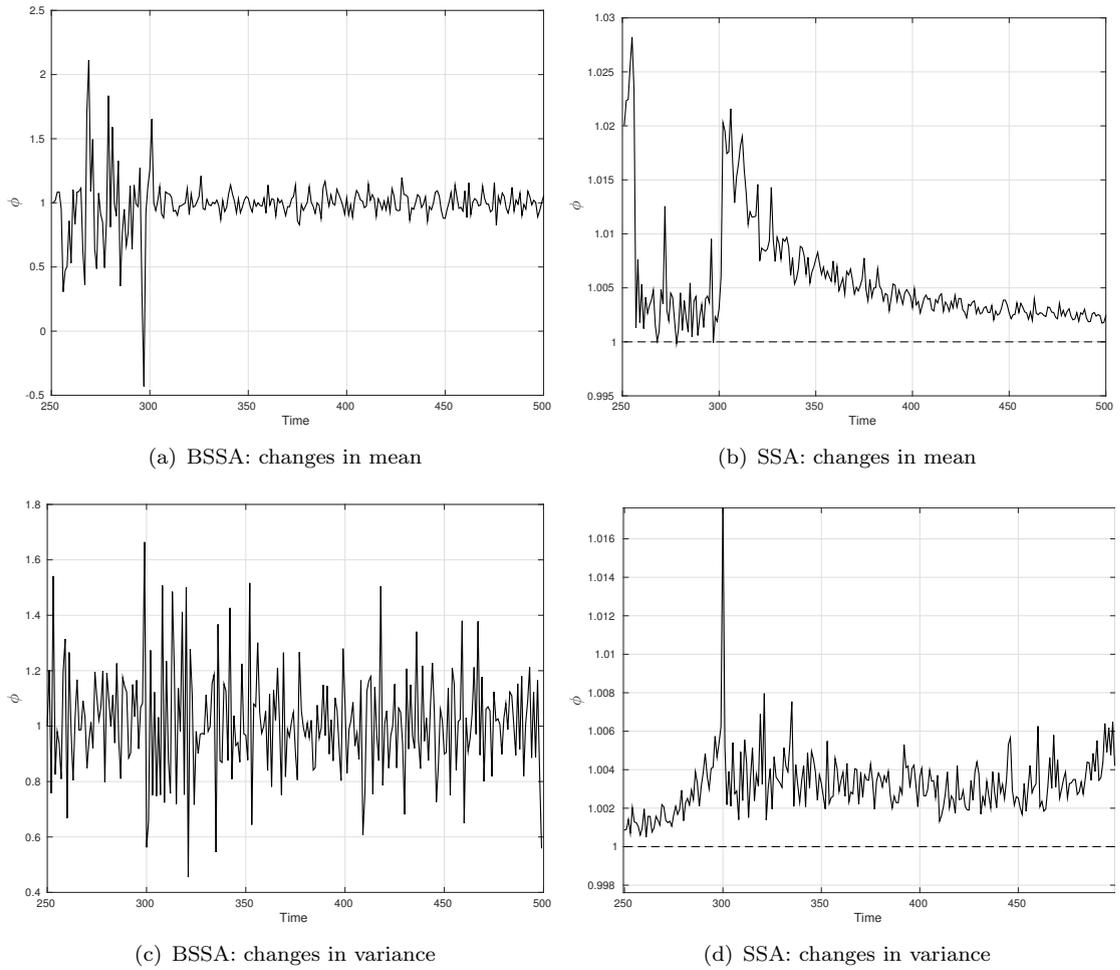


FIGURE 4.2: Coefficients of the LRF for univariate synthetic data when $L = 2$.

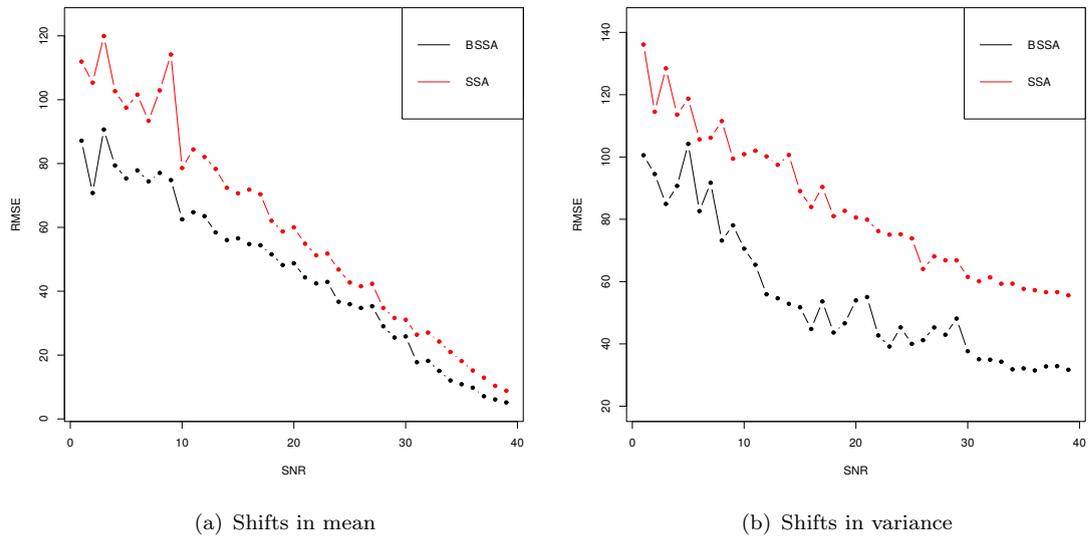


FIGURE 4.3: Effect of noise level on SSA and BSSA for the synthetic data.

4.3.2 Real data

4.3.2.1 Data, descriptive statistics and structural Breaks

The data used in this section are taken from I.N.S.E.E (Institute National de la Statistique et des Etudes Economiques) for France, from Statistisches Bundesamt, Wiesbaden for Germany and from the Office for National Statistics (ONS) for the UK and represent eight major components of real industrial production in France, Germany and the UK. These are seasonally adjusted series which measure real output for all facilities located in the manufacturing of Electric and Gas (E&G) utilities, Chemicals, Fabricated Metals (F.M.), Vehicle, Food Products (F.P.), Basic Metals (B.M.), Electrical Machinery (E.M.) and Machinery. These indicators show movements in production output and highlight structural developments as well as short-term changes in the economy. Although we consider only those eight of the two digit industries in this study, these eight industries account for more than 50% of the total industrial production in each country. The same eight industries have been considered by (Heravi et al., 2004), (Hassani et al., 2009) and (Osborn et al., 1999).

In all cases our sample period ends in February 2014. However, the data for France start from January 1990, for Germany from January 1991 and for the UK start from January 1998. Figures A.1, A.2 and A.3 show the series used in this study. Periods of overall expansion and contraction are evident in the graphs. As can be seen, most series for France and Germany present a long period of growth in 1990's and up to the current recession of 2008-2009. For the UK, however, most series show a period of stagnation in early 2000's and recession in 2008. For almost all the series, the steep drop in production can be seen around 2008-2009, which is attributed to the banking crisis and current recession.

TABLE 4.1: Descriptive statistics of Industrial Production Indicators for France.

	Weight	N	Mean In-sample	Mean Out-Of-Sample	Mean Overall	SD	SW	Break Points (Out-Of-Sample)
E&G	10.20	290	0.29	0.04	0.22	4.46	0.00*	May-09 (134)
Chemical	8.50	290	0.09	0.42	0.13	2.46	0.00*	May-10 (43)
F.M.	3.80	290	0.08	-0.34	-0.03	2.36	0.00*	Nov-08 (107)
Vehicle	5.80	290	0.29	-0.27	0.13	5.06	0.00*	Apr-08 (108)
F.P.	7.50	290	0.06	0.02	0.04	1.65	0.11	Jan-08 (44)
B.M.	3.00	290	0.06	-0.35	-0.05	2.07	0.00*	Nov-08 (161)
E.M.	4.70	290	0.10	-0.28	0.01	2.06	0.00*	Nov-08 (105)
Machinery	6.70	290	0.11	-0.41	-0.02	1.93	0.00*	Nov-08 (105)

Note:* indicates results are statistically significant at $\alpha=0.01$ via Shapiro-Wilk test.

TABLE 4.2: Descriptive statistics of Industrial Production Indicators for Germany.

	Weight	N	Mean	Mean	Mean	SD	SW	Break Points (Out-Of-Sample)
			In-sample	Out-Of-Sample	Overall			
E&G	9.00	278	0.06	0.01	0.04	2.83	0.02	Nov-02 (134)
Chemical	8.90	278	0.15	0.09	0.16	2.18	0.00*	Jul-10 (112)
F.M.	4.30	278	0.11	0.50	0.18	2.01	0.00*	Oct-08 (101)
Vehicle	9.80	278	0.24	0.64	0.33	4.93	0.00*	Sep-08 (124)
F.P.	8.60	278	0.16	-0.04	0.13	1.94	0.53	Jun-10 (106)
B.M.	3.90	278	0.03	0.30	0.07	3.20	0.00*	Oct-08 (104)
E.M.	7.10	278	0.20	0.43	0.24	2.07	0.00*	Dec-08 (101)
Machinery	9.60	278	0.02	0.52	0.11	2.97	0.00*	Dec-08 (99)

Note:* indicates results are statistically significant at $\alpha=0.01$ via Shapiro-Wilk test.

TABLE 4.3: Descriptive statistics of Industrial Production Indicators for the UK.

	Weight	N	Mean	Mean	Mean	SD	SW	Break Points (Out-Of-Sample)
			In-sample	Out-Of-Sample	Overall			
E&G	7.60	194	0.18	-0.17	0.09	3.25	0.00*	Oct-08 (64)
Chemical	8.60	194	-0.01	-0.17	-0.05	1.91	0.01	Oct-08 (64)
F.M.	4.50	194	-0.10	-0.06	-0.06	2.12	0.00*	Oct-08 (64)
Vehicle	7.20	194	0.10	0.37	0.22	5.72	0.00*	Aug-08 (66)
F.P.	13.60	194	-0.02	0.05	0.01	1.36	0.60	Jan-08 (73)
B.M.	5.60	194	-0.22	0.06	-0.08	3.86	0.00*	Oct-09 (64)
E.M.	10.40	194	-0.21	-0.16	-0.09	3.22	0.00*	Dec-09 (62)
Machinery	6.50	194	0.07	0.00	0.06	2.73	0.00*	Apr-09 (70)

Note:* indicates results are statistically significant at $\alpha=0.01$ via Shapiro-Wilk test.

Tables 4.1, 4.2 and 4.3 also show the Bai and Perron test statistics (Bai & Perron, 2003) for detecting structural breaks in these time series. Almost all the series show one or multiple breaks over the time period. However, only the date of the last break point is reported for each series in the tables. Results for all the three countries, based on the Bai and Perron test, indicate that all sectors are affected by the current recession of 2008-2009, except for E&G consumption for Germany. The results also indicate the break points for food and chemicals in June and July 2010 for Germany and May 2010 for food in France.

Tables 4.1, 4.2 and 4.3 also show the descriptive statistics for the monthly percentage changes in the original series, i.e. $100\left(\frac{y_t - y_{t-1}}{y_{t-1}}\right)$. In addition to reporting the growth/decline for the whole period, we also report the monthly percentage changes before and after the break points for each series. Overall all sectors have experienced growth over the whole period, with the exception of B.M., fabricated metals and machinery for France and the UK. Moreover, for the UK, chemicals also show a decline over this period. Some sectors have experienced substantial growth in production over 1990's and early 2000's

for Germany and France. In particular, electricity/gas and vehicle production in France show average increase of around 0.30 percentage per month or 3.6% per year. The results for Germany show that all sectors have recovered after the recession, with the exception of F.P., and in particular vehicles prediction shows an average increase of 0.64 percentage or about 8% growth per year. Declining industries after the break points are mostly in France and the UK, with machinery showing the highest average decline of 0.41% per month.

TABLE 4.4: Univariate post-sample forecast accuracy measures for France.

Series	Steps	RMSE						
		SSA	Boot SSA	BSSA	ARIMA	ARFIMA	ETS	GARCH
E&G	h=1	3.76	3.76	2.72 *	3.88	3.60	3.73	3.63
	h=3	4.45	4.46	2.85 *	4.25	4.19	4.08 *	4.19 *
	h=6	4.96	4.98	3.51 *	4.69	4.72	4.40 *	4.60 *
	h=12	5.61	5.64	4.69 *	5.42	5.07	4.72 *	5.15 *
Chemical	h=1	1.99	1.99	1.50 *	2.08	2.10	4.42	2.02
	h=3	3.01	3.01	2.41 *	3.08	3.18	3.07	3.07
	h=6	3.94	3.94	3.30 *	3.94	4.18	4.00	3.86
	h=12	3.45	3.45	3.19 *	3.40	4.04	3.49	3.27
F.M.	h=1	2.99	2.99	1.52 *	2.99	2.94	3.20	2.08
	h=3	5.50	5.50	1.89 *	5.21	5.10	5.35	2.62 *
	h=6	8.73	8.73	4.14 *	8.57	8.27	8.59	3.56 *
	h=12	13.74	13.74	12.80	13.20	12.55 *	13.72	13.02 *
Vehicle	h=1	6.50	6.50	3.89 *	6.79	6.81	3.21	4.37
	h=3	10.34	10.35	4.80 *	10.91	10.04	5.35	5.54
	h=6	15.50	15.53	7.64 *	16.36	14.76	8.60	6.86 *
	h=12	21.69	21.74	14.97	22.84	19.59 *	13.72 *	18.42 *
F.P.	h=1	1.27	1.27	1.02 *	1.07 *	1.28	5.07	1.77
	h=3	1.35	1.35	1.02 *	1.19 *	1.46	1.44	2.45
	h=6	1.84	1.84	1.41 *	1.52 *	1.87	1.85	3.39
	h=12	2.54	2.53	2.36 *	1.90 *	2.34	2.42	5.06
B.M.	h=1	2.73	2.73	1.48 *	2.62	2.50 *	3.05	4.65 *
	h=3	4.98	4.98	1.84 *	4.82	4.48 *	5.18	7.84
	h=6	7.78	7.78	3.75 *	8.14	7.35	8.43	9.31
	h=12	11.71	11.71	10.82	13.32	11.18 *	13.31	11.23
E.M.	h=1	2.98	2.98	1.50 *	2.95	2.78	2.82	4.74
	h=3	5.14	5.14	1.99 *	4.96 *	4.77	4.45 *	6.53
	h=6	8.33	8.35	4.83 *	8.10 *	8.15	7.40 *	9.15
	h=12	13.75	13.76	12.71	13.19 *	13.41	13.18	13.75
Machinery	h=1	3.89	3.90	1.44 *	3.30	3.32 *	3.33	3.35
	h=3	7.02	7.03	1.83 *	5.77 *	5.76 *	5.75 *	6.20
	h=6	11.58	11.59	5.42 *	9.94 *	10.01 *	9.93 *	9.87
	h=12	19.47	19.49	17.71 *	17.93 *	18.09 *	18.25	21.39
Summary	h=1		0	8	1	2	0	1
	h=3		0	8	3	2	3	2
	h=6		0	8	3	1	3	3
	h=12		0	4	3	4	2	3

Note: * indicates results are statistically significant at $\alpha=0.01$ based on modified GN test.

TABLE 4.5: Univariate post-sample forecast accuracy measures for Germany.

Series	Steps		RMSE					
	h	SSA	Boot SSA	BSSA	ARIMA	ARFIMA	ETS	GARCH
E&G	h=1	3.47	3.47	1.79 *	3.22	3.22 *	3.42	3.27
	h=3	4.64	4.64	3.64 *	4.27 *	4.28 *	4.32 *	4.35
	h=6	5.38	5.38	4.52 *	4.71 *	4.73 *	4.85 *	4.87 *
	h=12	5.72	5.70	5.24 *	4.66 *	4.85 *	4.96 *	5.09 *
Chemical	h=1	2.31	2.31	1.25 *	2.30	2.24	2.29	2.27
	h=3	3.61	3.61	2.69 *	3.47	3.64	3.60	3.57
	h=6	5.34	5.35	4.20 *	4.80 *	5.42	5.27	5.54
	h=12	7.50	7.52	6.70 *	6.20 *	7.65	7.24	8.22
F.M.	h=1	2.49	2.49	1.02 *	2.97	2.30	2.65	2.52
	h=3	5.04	5.05	3.59 *	4.52 *	5.19	5.13	5.14
	h=6	8.24	8.26	6.89 *	7.53 *	8.90	8.22	8.60
	h=12	13.14	13.16	11.41 *	11.87 *	13.62	12.68	14.27
Vehicle	h=1	4.93	4.93	2.72 *	4.93	5.10	5.02	5.06
	h=3	7.20	7.21	4.59 *	7.17	7.31	7.20	7.19
	h=6	10.00	10.01	8.82	10.00	10.40	9.93	9.99
	h=12	14.21	14.23	13.26 *	13.36 *	14.89	13.85 *	14.58
F.P.	h=1	1.68	1.68	0.97 *	1.49 *	1.83	1.54 *	1.56 *
	h=3	1.83	1.84	1.55 *	1.65 *	2.10	1.68 *	1.76
	h=6	1.95	1.96	1.41 *	1.86	2.50	1.87	2.09
	h=12	2.46	2.48	2.08 *	2.29	3.68	2.23 *	2.90
B.M.	h=1	3.78	3.78	1.69 *	3.75	3.38 *	3.84	3.44 *
	h=3	7.06	7.07	5.41 *	6.69 *	6.27 *	7.03	6.46
	h=6	11.16	11.17	10.30	10.62 *	10.08 *	10.94 *	10.68
	h=12	15.59	15.60	14.88	14.42 *	13.11 *	15.01 *	14.34 *
E.M.	h=1	2.58	2.59	1.16 *	2.21 *	2.43	2.31	2.41
	h=3	4.95	4.96	3.16 *	4.00 *	5.29	4.15 *	4.47
	h=6	8.32	8.34	5.96 *	7.30 *	10.06	7.29 *	8.10
	h=12	13.75	13.78	11.34 *	11.77 *	15.78	12.81	16.26
Machinery	h=1	3.71	3.72	1.98 *	3.93	3.69	3.61	3.66
	h=3	5.81	5.82	3.78 *	5.83	5.76	5.20 *	5.22 *
	h=6	9.60	9.62	6.41 *	10.22	10.00	8.54 *	8.65 *
	h=12	16.11	16.14	14.09 *	19.62	17.22	15.27 *	15.95
Summary	h=1		0	8	2	2	1	2
	h=3		0	8	5	2	4	1
	h=6		0	6	5	2	4	2
	h=12		0	6	6	2	5	2

Note:* indicates results are statistically significant at $\alpha=0.01$ based on modified GN test.

TABLE 4.6: Univariate post-sample forecast accuracy measures for the UK.

Series	Steps		RMSE					
	h	SSA	Boot SSA	BSSA	ARIMA	ARFIMA	ETS	GARCH
E&G	h=1	3.19	3.20	1.69 *	3.23	3.07	3.26	3.20
	h=3	4.48	4.48	2.78 *	4.26	3.98 *	3.94 *	4.25
	h=6	4.62	4.62	3.73 *	5.06	4.09 *	4.20 *	4.66
	h=12	6.03	6.03	5.28 *	6.97	4.97 *	4.91 *	5.77
Chemical	h=1	2.52	2.53	2.11 *	2.55	3.10	2.55	2.50
	h=3	3.86	3.87	3.02 *	3.39 *	7.67	3.79	3.72
	h=6	4.80	4.82	3.22 *	4.45 *	11.48	4.78	4.90
	h=12	5.79	5.81	4.82 *	5.91	10.88	5.77	6.37
F.M.	h=1	2.56	2.57	2.30 *	2.66	2.54	2.67	2.55
	h=3	3.97	3.98	3.31 *	4.01	4.02	4.02	4.06
	h=6	5.11	5.12	3.94 *	5.03	5.22	5.04	5.78
	h=12	6.53	6.53	5.09 *	5.87 *	6.30	5.87 *	7.84
Vehicle	h=1	6.16	6.17	2.41 *	7.32	7.00	8.28	7.84
	h=3	10.55	10.57	7.36 *	10.49	11.80	12.22	12.08
	h=6	13.94	13.97	11.15 *	13.78	14.57	14.57	14.21
	h=12	14.60	14.65	11.45 *	13.76 *	12.18 *	14.14	10.30 *
F.P.	h=1	1.73	1.73	1.43 *	1.65	1.69	1.70	1.67
	h=3	2.34	2.34	1.50 *	2.28	2.38	2.33	2.33
	h=6	3.13	3.13	2.05 *	3.02	2.96	3.10	3.05
	h=12	4.87	4.87	4.08 *	4.54 *	4.16 *	4.69 *	4.56 *
B.M.	h=1	5.70	5.70	4.14 *	5.28 *	5.46	5.45 *	5.58
	h=3	7.96	7.98	7.02 *	7.48 *	7.67	7.70 *	7.92
	h=6	10.11	10.13	6.98 *	9.14 *	9.60	9.45 *	9.60
	h=12	11.42	11.41	8.92 *	8.84 *	9.65 *	9.72 *	9.69 *
E.M.	h=1	4.66	4.66	3.74 *	5.00	4.60	4.55	4.74
	h=3	6.39	6.40	4.59 *	6.52	6.17	6.19	6.53
	h=6	8.74	8.76	5.73 *	8.58	8.39	8.43	9.15
	h=12	11.11	11.15	9.12 *	10.24 *	9.57	10.29 *	11.47
Machinery	h=1	4.14	4.14	3.29 *	4.10	3.87	4.18	3.35
	h=3	6.16	6.17	3.38 *	6.25	5.66 *	6.43	6.20
	h=6	9.90	9.91	6.14 *	9.67	8.91 *	9.97	9.87 *
	h=12	15.15	15.16	12.16 *	14.35 *	12.91 *	14.60 *	15.23 *
Summary	h=1		0	8	1	0	0	0
	h=3		0	8	2	2	2	0
	h=6		0	8	2	2	2	1
	h=12		0	8	7	5	6	4

Note: * indicates results are statistically significant at $\alpha=0.01$ based on modified GN test.

The sample standard deviations indicate greater volatility for the vehicle series than those of other sectors and with very low volatility for F.P.. The results for normality test based on Shapiro-Wilk test also provide strong evidence of non-normality for all the series, except for the F.P.. The results are all statistically significant at 1% level except electricity/gas for Germany and F.P. for all the three countries.

4.3.2.2 Forecasting results

All comparisons are made in terms of the prediction RMSE with respect to univariate SSA for monthly IPI series as the benchmark. Tables 4.4, 4.5 and 4.6 display information on the RMSE for France, Germany and the UK. For assessing the statistical significance of forecasting methods we used Granger-Newbold (GN) test (Mizrach, 1996).

In this section we evaluate forecast performance of the Singular Spectrum Analysis with its Bayesian format (BSSA), bootstrap SSA and basic SSA. All models are estimated based on the data summarised in Tables 4.1, 4.2 and 4.3, as our interest is to assess the forecast accuracy in the presence of a structural break in the forecast period. Post-sample forecasts are then computed for the months after the break point to the end of the data, February 2014. Thus the number of observations retained for post-sample forecast accuracy test are different depending on the date of the break point in the series. However, as may be seen from the descriptive tables in Section 5, the number of observations retained for post-sample forecast accuracy evaluation are around 60 months, the minimum number of observations held are 43 and 44 months for chemicals and F.P. for Germany.

Forecast accuracy is measured based on the magnitude of forecast errors, such as the prediction Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). However, since these measures give quantitatively similar results and to conserve space, we only report the RMSE, as this is the most frequently quoted measure in forecasting (Zhang et al., 1998). Tables 4.4, 4.5 and 4.6 show the out-of-sample RMSE and the ratio of RMSE (RRMSE) results for France, Germany and the UK. Average RRMSE is also given for each horizon at the bottom of each table.

BSSA vs Boot SSA and SSA In order to obtain the average bootstrap forecasts (see 2.1.5.2) the procedure is replicated over 1000 times. The results show no evidence of any statistical difference between SSA and Bootstrap SSA, and in fact, they are very similar for all the horizons and all the three countries. Comparing BSSA with SSA, the results are statistically significant at %1 level for almost all the horizons and all the three countries. However, the quality of the forecast with BSSA is much better for $h = 1, 3$ and 6 and less significant for $h = 12$. The BSSA technique outperforms SSA and reduces the RMSE by 40% for France and Germany and 28% for the UK. The improvements for $h = 12$ are 10% for France and Germany and 18% for the UK.

BSSA vs ARIMA, ARFIMA, ETS and GARCH To better assess the forecast accuracy of BSSA, performance is compared with the most commonly used forecasting methods, such as ARIMA, ARFIMA and ETS. To determine the models, estimating the parameters and finally making a forecast an automatic version of these models is designed by (Hyndman & Khandakar, 2008) which uses a step-wise algorithm for forecasting and can be run to both seasonal and non-seasonal data.² The performance of the proposed model is also compared with that of a GARCH model. GARCH models can capture time variation in the full density parameters, with the AR conditional density model, by relaxing the assumption that the conditional distribution of the standardised innovations is independent of the conditioning information³.

As can be seen from Tables 4.4, 4.5 and 4.6, there is a clear indication that BSSA is most likely to provide more accurate forecasts than ARIMA, AFRIMA, ETS and GARCH. To be more specific, 100% of BSSA forecasts are statistically significant at horizons $h = 1$ and 3, 92% of for $h = 6$ and 75% of them for the longer horizon ($h = 12$) which is comparatively higher than those provided by ARIMA, ARFIMA, GARCH and ETS. It seems more likely latter models provide more significant forecasts for longer horizons than short or medium horizons. One reason could be that autoregressive models can be more adaptable after sudden changes to predict longer horizons.

Empirical cumulative distribution function Figure 4.4 presents the cumulative distribution function (c.d.f) of the RMSE values of the absolute values of the out-of-sample errors obtained by SSA, Bootstrap SSA and BSSA for all 24 time series. If the c.d.f. produced by one method is strictly above the c.d.f. obtained by another method, we may then say that the forecast errors are stochastically smaller for the first method. Figures 4.4(a), 4.4(b), 4.4(c) and 4.4(d) demonstrate that the forecast errors obtained by the BSSA are much smaller than the errors of the other two methods for $h = 1, 3, 6$ and 12, confirming the superiority of BSSA.

4.3.3 Summary and conclusion on univariate case

This section compared performance of Bayesian SSA with basic SSA and bootstrap SSA for forecasting synthetic models and industrial production indicators in France, Germany and the UK. It is found empirical evidence that the modified SSA technique with the Bayesian recurrent formula performs substantially better than SSA and Bootstrap SSA methods in the presence of a structural break, according to prediction root mean square

²The models can vary by their complexity and sensitivity to the data structure. An optimal version of ARIMA, ARFIMA and ETS models are provided through a forecasting package in R, available from CRAN at <https://cran.r-project.org/web/packages/forecast/index.html>. For a detailed description of the algorithm on which it is based, see (Hyndman & Khandakar, 2008).

³ The rugarch package for fitting a univariate GARCH model is available from the CRAN at [andhttps://cran.r-project.org/web/packages/rugarch/index.html](https://cran.r-project.org/web/packages/rugarch/index.html).

TABLE 4.7: Summary statistics for univariate out-of-sample forecasting accuracy measures for Industrial Production Indicators for all three countries.

Steps	RMSE		RRMSE		Sig. at	
	SSA	Boot SSA	BSSA	$\frac{BootSSA}{SSA}$		$\frac{BSSA}{SSA}$
h=1	3.406	3.408	2.125	1.001	0.642	22
h=3	5.318	5.326	3.386	1.002	0.656	14
h=6	7.623	7.637	5.277	1.002	0.706	12
h=12	10.664	10.679	9.364	1.001	0.876	9
Overall	6.753	6.762	5.038	1.001	0.720	57

error. Table 4.7 presents the summary statistics of the RMSE, RRMSE and the number of significant forecasts predicted across all series and countries. The results indicate the superiority of using the Bayesian LRF for out-of-sample forecasting, with overall reduction of 28% according to RMSE criterion. The results also show that the improvement is 36% for a one step-ahead forecast, $h = 1$, decreasing to 13% as h increases to 12 months ahead.

Comparing BSSA forecasts with SSA, BSSA outperforms SSA significantly in 22, 14, 12, and 9 times out of 24 cases) at $h = 1, 3, 6$ and 12 horizons respectively at 1% level. The last column in Table 4.7 shows the number of statistically significant cases and indicates that for all the horizons and across all three countries, BSSA outperforms SSA significantly at 1% level in 60% of cases (57 out of 96 cases). The graph of the cumulative density function also confirms the findings, showing that the errors obtained by the BSSA are stochastically smaller than the errors obtained by the other models for $h = 1, 3$ and 6. Next we shall investigate the performance of BSSA in a multivariate setting.

4.4 Empirical results; Multivariate

In order to evaluate the experimental results for the multivariate models, a number of experiments are conducted again using both synthetic and real data. Both MSSA (Hasani & Mahmoudvand, 2013) and MBSSA are applied to synthetic time series with shifts and structural breaks and the results are discussed in the next section. The second part looks at real data and presents the resulting predictions.

4.4.1 Synthetic data

To simulate multivariate time series, the popular Vector AutoRegressive (VAR) model is used which has been found in several studies to model dynamic behaviour in financial

and economic time series (Lütkepohl, 2005). Let $Y = (y_k^{(1)}, \dots, y_k^{(M)})$ denote a vector of time series, and the basic p -lag, VAR(p) model, has the form:

$$Y_k = \delta + \mathbf{\Pi}_1 Y_{k-1} + \mathbf{\Pi}_2 Y_{k-2} + \dots + \mathbf{\Pi}_p Y_{k-p} + \epsilon_k \quad (4.19)$$

where $\mathbf{\Pi}_i \in \mathcal{R}^{M \times M}$ is a matrix of coefficients, $\epsilon_k \in \mathcal{R}^{M \times 1}$ is white noise vector process with covariance matrix Σ , and δ is an intercept.

The simulation seeks to give insight about the estimated coefficients of the LRF in the presence of a shift in the mean, and thus multivariate time series are generated using (4.19) in which a shift in the mean is artificially introduced at time 150. Figures 4.5(a), 4.5(d) and 4.6(a) show one realisation of each model with corresponding parameter estimates. Next, the performance of MSSA and MBSSA following the structural break is examined, *i.e.* from time 150 to 250, over 500 realisations of the process.

The first experiment deals with two mean-inflated series. Figures 4.5(b) and 4.5(c) illustrate the leading coefficients of LRF estimated using MSSA and MBSSA. Following the structural break, the parameters estimated by MSSA rise rapidly from 1.005 to 1.03 and decay slowly to their pre-structural break values. In contrast, coefficients estimated by MBSSA also grow significantly from 1.00 to 1.20 but thereafter decay *exponentially* as we move away from the shift. The response time for MBSSA (10% to 90% of its final value) is 20 steps while for MSSA this is far greater and is 70 steps. In the second experiment only one of the time series is affected by a mean shift, and it is observed that only the coefficients corresponding to that time series are affected under MBSSA (Figure 4.5(e)). The third simulation consists of 6 series with different levels of mean, non-affected, moderately-inflated and highly-inflated. Based on Figure 4.6(b), coefficients estimated by MSSA decay exponentially over time whereas MBSSA could accurately find a robust estimator of LRF parameters by tracking their movement over the state of its related series. Technically, the presence of a constant term, δ , in equation (4.19) can be related to an eigenvector of $[1, 1, \dots, 1]^T 1/\sqrt{L}$ in U_i (see Theorem 4.1). Consequently, the empirical coefficients of LRF demonstrate an *exponential* decay in which they die out slowly as we move away from the shift and it can be proportional to their linear continuation of the time series pre-break.

4.4.2 Real data

Empirical evidence suggests that the *shared* dynamical pattern in the Industrial Production Indicator (IPI) time series has strengthened over the years and should be taken into account during modelling and forecasting, see (Groth & Ghil, 2011; Heravi et al., 2004; Osborn et al., 1999). These changes can bring more instability into the model in terms of parameter estimation and even model selection as noted by (Groth & Ghil, 2011).

TABLE 4.8: Variance Inflation Factor to test multi-collinearity.

Series	France	Germany	UK
E&G	6.71*	2.13	2.10
Chemical	5.21*	40.57*	6.04*
F.M.	71.94**	53.24**	17.45**
Vehicle	10.28**	18.41**	2.15
F.P.	2.77	28.59**	1.69
B.M.	80.17**	5.81*	9.13*
E.M.	27.28**	47.64**	4.78
Machinery	13.45**	22.06**	1.80

** and *: show high and moderate level of multi-collinearity.

TABLE 4.9: P-values of Augmented Dickey-Fuller test.

Series	France	Germany	UK
E&G	0.41**	0.34**	0.56**
Chemical	0.27**	0.04*	0.38**
F.M.	0.56**	0.01	0.49**
Vehicle	0.89**	0.12**	0.35**
F.P.	0.38**	0.76**	0.34**
B.M.	0.56**	0.01	0.11**
E.M.	0.41**	0.03*	0.25**
Machinery	0.39**	0.01	0.04*

** and *: Indicate results are statistically significant at $\alpha = 0.01$ and $\alpha = 0.05$.

4.4.2.1 Variance Inflation Factor

To test the shared dynamics hypothesis on IPI time series, the variance inflation factor (VIF⁴) as a measure of collinearity of multiple time series is used. Based on Table 4.8, almost half of the series are shown to be strongly correlated and a quarter of them are moderately correlated. This can be mainly attributed to France and German's economy which has a larger industrial component than the UK's. A large value for the VIF tells us that the variance of the estimated coefficient of corresponding series is inflated significantly because it is strongly correlated with at least one of the series.

On the other hand, the Augmented Dickey-Fuller test statistic in table 4.9 indicate that the majority of these time series are non-stationary, which is common for economic time series.

⁴ VIF regresses an explanatory variable toward the rest of variables and subsequently measures the square of the multiple correlation coefficients by $VIF = \frac{1}{1-R^2}$. A rule of thumb is that if VIF goes above 10, then multi-collinearity is high while for $5 < VIF < 10$ multi-collinearity is moderate.

TABLE 4.10: Engle-Granger cointegration test.

Series	Series	France	Germany	UK
E&G	Chemical	0.01 **	0.01 **	0.04 **
E&G	F.M.	0.01 **	0.01 **	0.09
E&G	Vehicle	0.01 **	0.01 **	0.09
E&G	F.P.	0.01 **	0.01 **	0.10
E&G	B.M.	0.01 **	0.01 **	0.06
E&G	E.M.	0.01 **	0.01 **	0.08
E&G	Machinery	0.01 **	0.01 **	0.09
Chemical	F.M.	0.03*	0.05	0.09
Chemical	Vehicle	0.09	0.01 **	0.59
Chemical	F.P.	0.01 **	0.01 **	0.50
Chemical	B.M.	0.02 *	0.03 *	0.38
Chemical	E.M.	0.02*	0.06	0.33
Chemical	Machinery	0.03 *	0.03 *	0.53
F.M.	Vehicle	0.33	0.01 **	0.39
F.M.	F.P.	0.80	0.21	0.63
F.M.	B.M.	0.01 **	0.02*	0.29
F.M.	E.M.	0.02	0.21	0.32
F.M.	Machinery	0.20	0.01 **	0.54
Vehicle	F.P.	0.60	0.14	0.19
Vehicle	B.M.	0.34	0.02*	0.22
Vehicle	E.M.	0.20	0.27	0.40
Vehicle	Machinery	0.45	0.10	0.36
F.P.	B.M.	0.01 **	0.05	0.07
F.P.	E.M.	0.01 **	0.10	0.07
F.P.	Machinery	0.01 **	0.17	0.01**
B.M.	E.M.	0.02 *	0.14	0.04*
B.M.	Machinery	0.23	0.05	0.24
E.M.	Machinery	0.18	0.02*	0.21

** and *: Indicate results are statistically significant at $\alpha = 0.01$.

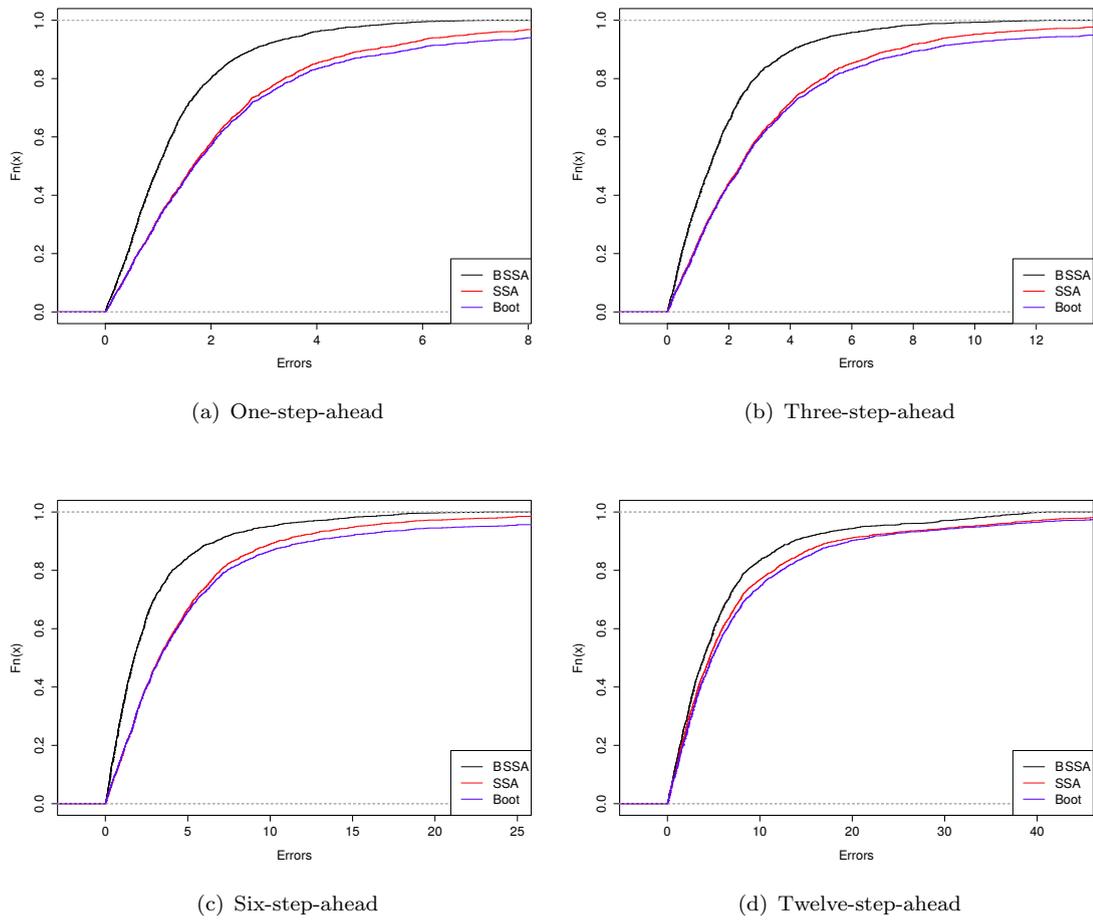


FIGURE 4.4: Empirical cumulative distribution functions of absolute values of forecast errors for SSA, Bootstrap SSA and BSSA.

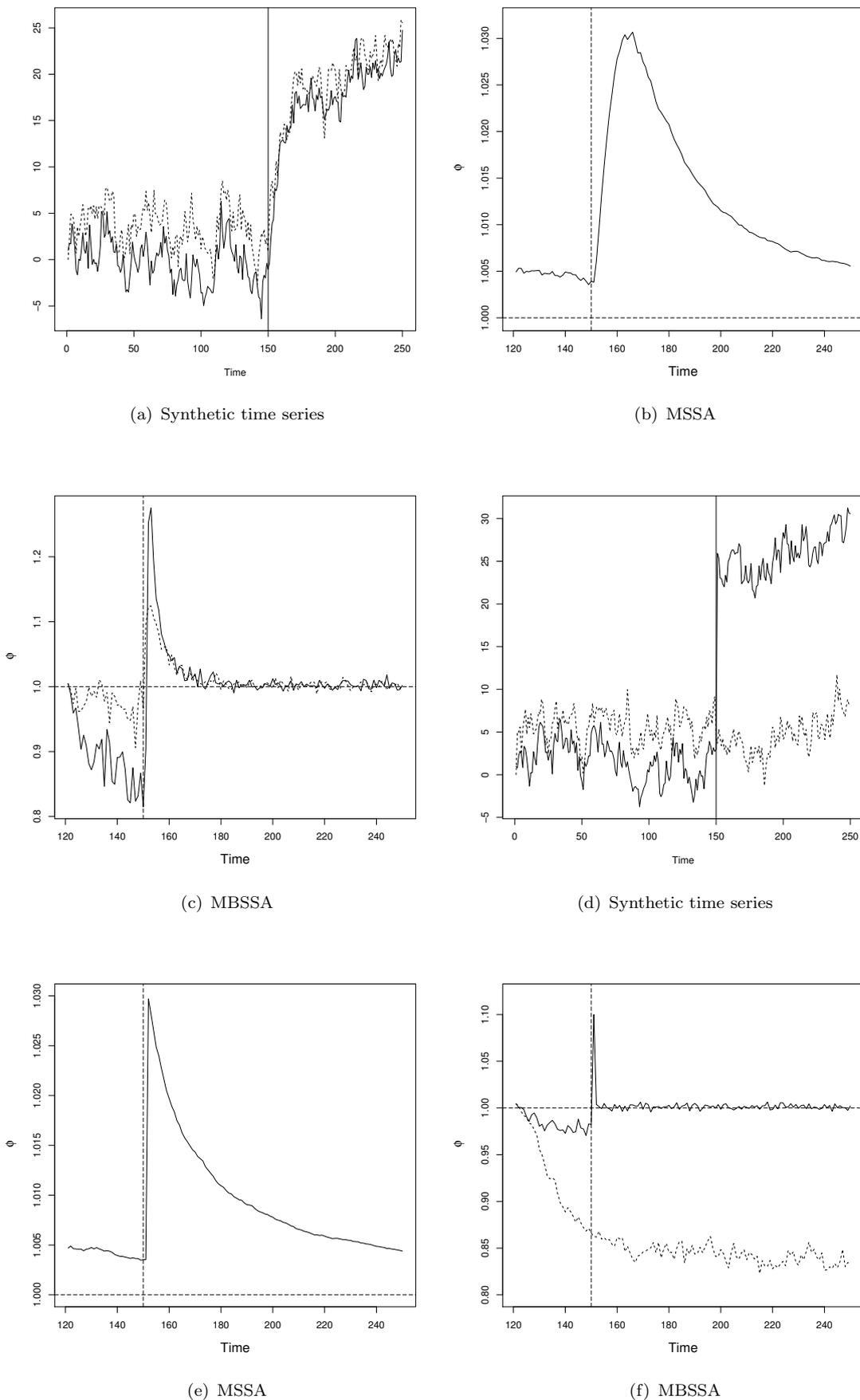
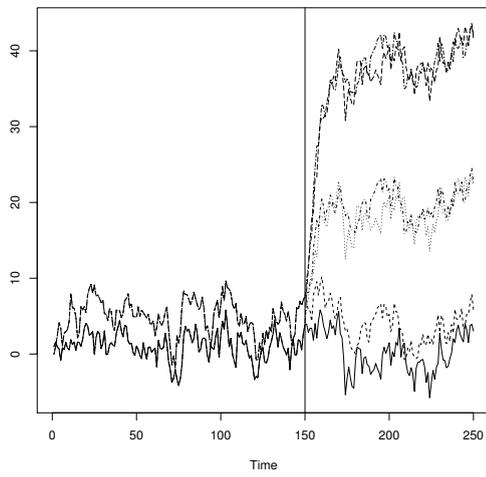
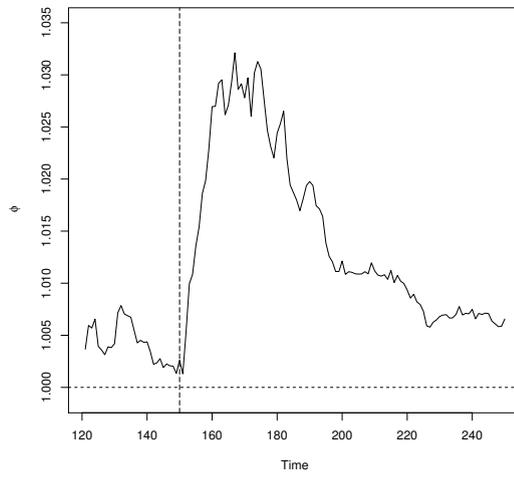


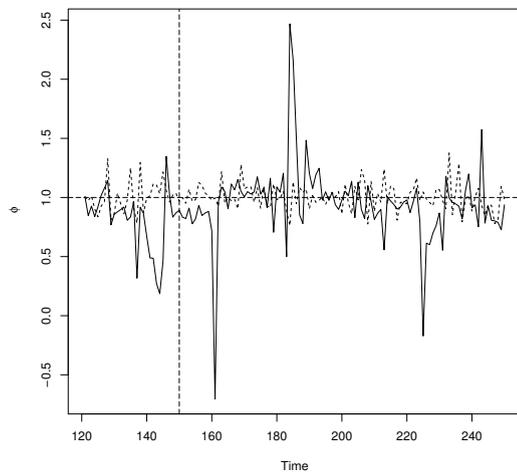
FIGURE 4.5: Coefficients of the LRF for bivariate synthetic data ($L = p = 2$), $\delta = (-0.7, 1.3)$, $\delta_s = (1.76, 2.43)$, $\Pi_1 = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}$.



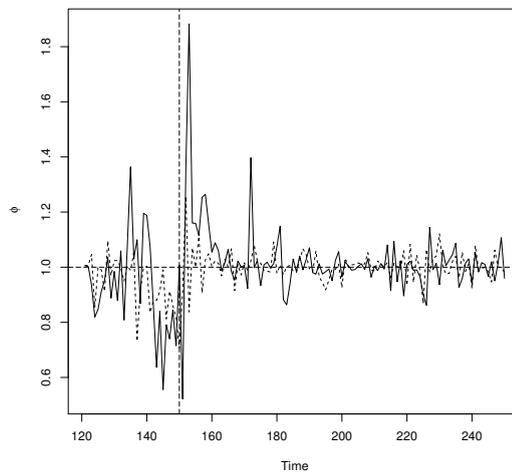
(a) Synthetic time series



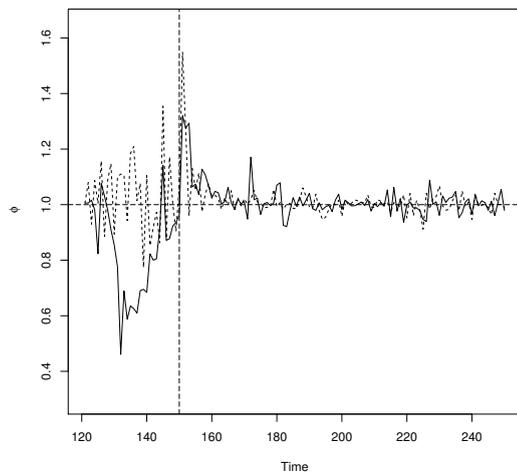
(b) MSSA



(c) MBSSA (non-affected series)



(d) MBSSA (moderately mean-inflated series)



(e) MBSSA (highly mean-inflated series)

FIGURE 4.6: Coefficients of the LRF for $(L = p = 2, \delta = (-0.7, 1.3), \delta_{s_1} = (1.76, 2.43),$

$$\delta_{s_2} = (2.64, 3.73), \mathbf{\Pi}_1 = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}.$$

TABLE 4.11: Multivariate post-sample forecast accuracy measures for France.

Series	Series	Steps	RMSE								
			h	SSA	MSSA-2	MBSSA-2	VAR	VECM	MSSA-8	MBSSA-8	ICA
										MSSA-8	MBSSA-8
E&G	E.M.	h=1	3.76	3.75	2.70 *	3.85	3.70	3.74	2.88 *	2.32 *	2.23 *
	Chemical	h=3	4.45	4.38 *	3.32 *	4.30	4.25	4.37 *	3.34 *	2.61 *	2.40 *
	B.M.	h=6	4.96	4.80 *	4.03 *	4.62 *	4.62	4.78 *	4.71	3.39 *	3.09 *
	B.M.	h=12	5.61	5.23 *	4.93 *	5.91 *	4.99	5.21 *	5.00 *	4.49 *	4.19 *
Chemical	Machinery	h=1	1.99	1.98	1.44 *	1.94	1.86	2.00	1.53 *	1.77	1.23 *
	Machinery	h=3	3.01	2.98	2.08 *	3.01	2.78	3.05	2.24 *	1.99	1.83 *
	E&G	h=6	3.94	3.87 *	3.32 *	3.57 *	3.59	4.03	3.48	2.42 *	1.67 *
	E&G	h=12	3.45	3.36	3.58	3.20 *	2.99 *	3.53	3.61	2.85 *	2.11 *
F.M.	Chemical	h=1	2.99	2.99	1.47 *	2.92	2.80	3.00	1.55 *	2.72	1.49 *
	Vehicle	h=3	5.50	5.49 *	2.08 *	5.62 *	5.20	5.53	2.16 *	5.2	2.26 *
	Vehicle	h=6	8.73	8.72	3.55 *	9.53 *	8.45	8.81	3.80 *	8.51	3.92 *
	Chemical	h=12	13.74	13.69 *	12.77	15.61 *	13.33	13.95	13.70	13.53	11.82 *
Vehicle	F.P.	h=1	6.50	6.48 *	4.01 *	7.27	6.44	6.47	4.01 *	5.44 *	3.34 *
	F.M.	h=3	10.34	10.26 *	5.42 *	11.99	10.15	10.24 *	5.82 *	9.47 *	4.10 *
	F.M.	h=6	15.50	15.29 *	10.02 *	19.08	15.14	15.25 *	12.43	14.93 *	8.74 *
	F.M.	h=12	25.69	25.12 *	23.45	28.03 *	26.60	21.03 *	20.34	22.36 *	18.85 *
F.P.	Vehicle	h=1	1.27	1.27	1.17	1.43	1.27	1.27	1.18	0.95 *	0.56 *
	Vehicle	h=3	1.35	1.33 *	1.23	1.66	1.42	1.33	1.26	1.07 *	0.79 *
	Vehicle	h=6	1.84	1.79	1.21 *	1.85	1.78	1.79	1.27 *	1.19 *	0.98 *
	B.M.	h=12	2.54	2.31 *	1.96 *	2.37 *	2.36	2.36 *	2.00 *	1.37 *	1.11 *
B.M.	Chemical	h=1	2.73	2.73	1.84 *	2.50	2.59	2.74	1.66 *	2.54 *	1.44 *
	Vehicle	h=3	4.98	4.98	3.58 *	4.95 *	4.74	5.02	2.11 *	4.56	1.95 *
	F.P.	h=6	7.78	7.77	5.63 *	8.31 *	7.59	7.88	3.77 *	7.30 *	3.36 *
	Chemical	h=12	11.71	11.67 *	12.14	13.15 *	11.42	11.98	11.89	11.43	11.04 *
E.M.	E&G	h=1	2.98	2.96	1.59	2.61	2.95	2.97	1.52	2.27	1.52 *
	Vehicle	h=3	5.14	5.09	3.25	4.35	4.86	5.11	2.42	4.30	1.97 *
	Chemical	h=6	8.33	8.26	5.07	7.82	7.96	8.27	3.60	6.95 *	3.54 *
	E&G	h=12	13.75	13.57	11.26	14.39	13.10	13.62	9.98	10.87 *	9.98 *
Machinery	Chemical	h=1	3.89	3.88	1.86	3.46	3.55	3.88	1.41	2.84	1.41 *
	F.M.	h=3	7.02	6.97	3.77	6.43	6.32	6.97	2.94	5.46	2.03 *
	B.M.	h=6	11.58	11.47	5.56	11.32 *	10.80	11.46	5.09	8.88 *	4.19 *
	F.P.	h=12	19.47	19.18	14.74	20.39 *	18.35	19.15	12.54	13.96 *	12.54 *
Summary		h=1		1	5	0	0	0	5	3	8
		h=3		4	5	2	0	2	5	2	8
		h=6		3	6	5	0	2	3	6	8
		h=12		4	2	7	1	3	1	5	8

Note: * indicates results are statistically significant at $\alpha=0.01$ based on GN test.

TABLE 4.12: Multivariate post-sample forecast accuracy measures for Germany.

Series	Series	Steps	RMSE								
			h	SSA	MSSA-2	MBSSA-2	VAR	VECM	MSSA-8	MBSSA-8	ICA MSSA-8
E&G	Chemical	h=1	3.47	3.47	2.18	3.18	3.37	3.47	2.18 *	0.55 *	0.37 *
	Machinery	h=3	4.64	4.64	3.52 *	4.20 *	4.75	4.66	3.55 *	1.08 *	0.71 *
	Machinery	h=6	5.38	5.38	5.44	5.01	5.09	5.47	5.59	1.79 *	1.23 *
	Machinery	h=12	5.72	5.74	5.75	5.89	5.10 *	5.59 *	5.78	2.91 *	2.33 *
Chemical	E&G	h=1	2.31	2.31	1.47	2.19	2.21	2.31	1.47 *	1.33	0.89 *
	E&G	h=3	3.61	3.59	2.08 *	3.41	3.45	3.63	2.08 *	2.36	1.41 *
	E&G	h=6	5.34	5.28	3.94 *	5.07 *	5.18	5.37	3.94 *	3.80 *	2.67 *
	F.P.	h=12	7.50	7.31 *	6.50 *	7.32	7.20 *	7.58 *	7.39	6.21 *	5.34 *
F.M.	E&G	h=1	2.49	2.48 *	1.02	1.96 *	2.16	2.49	1.02 *	2.35	1.45 *
	E&G	h=3	5.04	5.02 *	2.27 *	4.35 *	4.69	5.03	2.27 *	4.58 *	2.91 *
	E&G	h=6	8.24	8.16	5.89 *	7.25 *	7.97	8.19	5.88 *	7.41 *	5.62 *
	E&G	h=12	13.14	12.83 *	11.97 *	11.47 *	12.77	12.95 *	11.99 *	11.60 *	10.57 *
Vehicle	E&G	h=1	4.93	4.92	2.85	5.15	4.81	4.92	2.85 *	2.89 *	1.79 *
	F.P.	h=3	7.20	7.17 *	3.86 *	7.68	7.37	7.19	4.86 *	5.45 *	3.30 *
	F.P.	h=6	10.00	9.92 *	8.37	10.97	10.59	9.98	9.74	8.77 *	6.97 *
	E&G	h=12	14.21	13.92 *	13.13 *	14.90	14.64	14.05	13.13 *	13.75	11.66 *
F.P.	B.M.	h=1	1.68	1.67	1.47	1.88	1.76	1.67	1.47 *	1.06 *	0.92 *
	E&G	h=3	1.83	1.80	1.40 *	2.26	1.88	1.82	1.40 *	1.22 *	0.86 *
	E&G	h=6	1.95	1.86 *	1.63 *	2.79	2.08	1.93	1.63 *	1.62 *	1.11 *
	E&G	h=12	2.46	2.19 *	1.63 *	4.05	2.53	2.46	1.87 *	2.47	1.91 *
B.M.	E&G	h=1	3.78	3.77 *	1.77 *	3.08	3.23 *	3.78	1.77 *	3.40 *	1.92 *
	F.P.	h=3	7.06	7.04 *	4.01 *	6.22	6.53	7.07	4.01 *	6.48 *	3.94 *
	E.M.	h=6	11.16	11.08 *	9.41	10.54	11.21	11.18	9.41	10.34 *	9.09 *
	Machinery	h=12	15.59	15.38 *	14.72 *	15.24	15.97	15.71	26.80	15.32	14.46 *
E.M.	E&G	h=1	2.58	2.56	1.10	2.16	2.21 *	2.57	1.10 *	2.65	1.50 *
	E&G	h=3	4.95	4.89 *	2.16 *	4.07 *	4.35 *	4.90	2.16 *	4.88	2.58 *
	E&G	h=6	8.32	8.13	4.96 *	7.35	7.80	8.14	4.96 *	7.87	5.68 *
	F.M.	h=12	13.75	13.08 *	10.99 *	12.75	12.97	13.15 *	13.32	12.59 *	10.96 *
Machinery	E&G	h=1	3.71	3.69 *	2.09	3.31	3.30 *	3.69 *	2.09 *	2.98 *	1.98 *
	F.M.	h=3	5.81	5.74 *	3.64 *	4.33 *	4.91 *	5.73 *	4.32 *	5.91	4.11 *
	F.P.	h=6	9.60	9.48 *	6.24 *	7.28 *	8.65 *	9.43 *	6.52 *	9.53	7.62 *
	E&G	h=12	16.11	15.73 *	13.41 *	13.38 *	14.90 *	15.73	13.69	14.63 *	13.88 *
Summary		h=1		3	1	1	3	1	8	5	8
		h=3		5	8	4	1	1	8	5	8
		h=6		4	5	3	1	1	5	6	8
		h=12		7	7	2	3	3	3	5	8

Note: * indicates results are statistically significant at $\alpha=0.01$ based on GN test.

TABLE 4.13: Multivariate post-sample forecast accuracy measures for the UK.

Series	Series	Steps	RMSE								
			h	SSA	MSSA-2	MBSSA-2	VAR	VECM	MSSA-8	MBSSA-8	ICA MSSA-8
E&G	B.M.	h=1	3.19	3.17	1.94 *	3.46	3.38	3.17	2.26 *	0.35 *	0.23 *
	Chemical	h=3	4.48	4.40	2.87 *	4.76	4.39	4.39	2.85 *	0.58 *	0.29 *
	F.M.	h=6	4.62	4.42	4.11	5.18	4.18	4.41	4.29	0.93 *	0.41 *
	F.P.	h=12	6.03	5.61 *	5.26 *	5.39	5.11 *	5.59	5.24 *	1.57 *	0.68 *
Chemical	Vehicle	h=1	2.52	2.50	1.71 *	2.49	2.56	2.50 *	1.80 *	1.81	0.74 *
	Vehicle	h=3	3.86	3.80	2.12 *	3.54	3.79	3.81 *	2.22 *	2.94	2.07 *
	E.M.	h=6	4.80	4.70	2.92 *	4.60	4.73	4.69 *	2.92 *	3.56 *	2.43 *
	F.M.	h=12	5.79	5.57 *	4.88 *	5.88	5.47	5.58 *	5.29 *	4.03 *	3.46 *
F.M.	Vehicle	h=1	2.56	2.55 *	1.78 *	2.75	2.63	2.56	2.20	2.54	0.99 *
	Vehicle	h=3	3.97	3.92 *	2.07 *	4.25	3.79	3.95	3.04 *	4.45	2.54 *
	Vehicle	h=6	5.11	5.00	2.76 *	5.21	4.88	5.04 *	3.80 *	5.18	3.71 *
	B.M.	h=12	6.53	6.09 *	4.22 *	5.80	6.45	6.37 *	4.97 *	6.21	5.64 *
Vehicle	B.M.	h=1	6.16	6.16	2.26 *	7.64	6.27	6.17	3.69 *	5.96	2.36 *
	E&G	h=3	10.55	10.53	5.69 *	13.63	10.40	10.57	5.66 *	10.47	5.77 *
	E&G	h=6	13.94	13.83	10.98	16.34	14.10	13.96	10.95 *	14.13	10.44 *
	E&G	h=12	14.60	13.99 *	15.02	15.47	14.34	14.60	14.97 *	15.87	11.12 *
F.P.	Machinery	h=1	1.73	1.73	1.18 *	1.82	1.72	1.74	1.30 *	0.65 *	0.49 *
	Machinery	h=3	2.34	2.31	1.43 *	2.81	2.28	2.36	1.54 *	1.05 *	0.63 *
	E.M.	h=6	3.13	3.06	2.00 *	3.80	3.08	3.20	2.02 *	1.71 *	1.10 *
	Machinery	h=12	4.87	4.66 *	3.61 *	5.03	4.65 *	5.10	3.96 *	2.71 *	1.84 *
B.M.	E&G	h=1	5.70	5.69	3.29 *	5.28	5.58	5.70	3.29 *	4.02 *	1.29 *
	E.M.	h=3	7.96	7.91	5.71 *	7.53	7.47	7.92	6.14	6.75 *	3.16 *
	E&G	h=6	10.11	9.91 *	6.45 *	9.84	9.81	9.86 *	8.19 *	7.88 *	5.46 *
	Machinery	h=12	11.42	10.58 *	10.96 *	10.45	10.78	10.41 *	14.57 *	8.46 *	7.42 *
E.M.	E&G	h=1	4.66	4.65	3.22 *	4.77	4.83	4.65	3.21 *	3.12 *	1.49 *
	B.M.	h=3	6.39	6.36	4.03 *	7.12	6.35	6.35	4.36 *	3.88 *	2.53 *
	Machinery	h=6	8.74	8.64 *	5.14 *	9.50	8.84	8.61 *	7.35	4.95 *	3.40 *
	Machinery	h=12	11.11	10.72 *	8.52 *	9.40	11.67	10.68 *	15.88 *	5.57 *	5.68 *
Machinery	F.P.	h=1	4.14	4.12	2.16 *	3.92	4.07	4.13	3.47 *	1.81 *	1.19 *
	Vehicle	h=3	6.16	6.11	3.01 *	6.22	6.08	6.11	4.67 *	3.06 *	1.70 *
	F.P.	h=6	9.90	9.77	5.04 *	9.60	9.63 *	9.76 *	7.59 *	4.82 *	3.38 *
	E.M.	h=12	15.15	14.86 *	11.87 *	14.37	14.51 *	14.85 *	13.27 *	6.07 *	4.57 *
Summary		h=1		1	8	0	0	1	7	6	8
		h=3		1	8	0	0	1	7	6	8
		h=6		2	6	0	1	5	6	6	8
		h=12		8	6	0	3	5	8	6	8

Note: * indicates results are statistically significant at $\alpha=0.01$ based on GN test.

4.4.2.2 Engle-Granger cointegration test

To measure how much these non-stationary time series are tied together the Engle-Granger cointegration test is applied (Table 4.10). It enables us to detect stable long-run relationships among non-stationary series with a tendency to revert toward a stochastic process.

The result shows that E&G is significantly cointegrated with the rest of the indices for both France and Germany at $\alpha = 0.01$. Even though the same conclusion can be drawn for Chemical in France (except Vehicle), in Germany, it is mainly cointegrated to Vehicle, F.P., B.M. and Machinery. F.P. is yet another indicator which depicts a long

run relationship with B.M. and E.M. and Machinery sector in French Economy (see Table 4.11). Similar considerations apply to F.M. and its cointegration with Vehicle and Machinery sectors in both countries. However, it should be pointed out that in the UK economy, “E&G and Chemical”, “F.P. and Machinery” and “B.M. and E.M” are the only cointegrated pairs detected by Engle-Granger test (Table 4.13).

4.4.2.3 Forecasting results

This section compares the performance of out-of-sample forecasts using MBSSA with the most commonly used models such as MSSA-2 (for any pairs), VAR, VECM⁵ and MSSA-8 (for all eight series together). On account of the presence of cointegration in the time series, we use Independent Component Analysis (ICA) as a pre-processing step before SSA due to the lack of strong separability (See Golyandina et al. (2001) on the use of ICA with SSA).

MSSA-2 vs MBSSA-2 Our analysis starts by applying MSSA and MBSSA on a joint combination of two time series⁶. Note that the optimal partner time series unexpectedly depends on the forecast horizon (for example E&G with E.M. for $h = 1$, but E&G with B.M. for $h = 6$, however the RMSE differences are slight). Tables 4.11, 4.12 and 4.13 show that multivariate models are superior to the univariate SSA models. For instance, in the majority of cases, MSSA performs better than SSA for horizons of up to a year ($\alpha = 0.01$). However, MBSSA-2 is generally superior to both. The statistical significance of the RMSEs (relative to the univariate RMSE), is also reported which shows that two thirds of French and nearly a quarter of forecast improvements (all horizons) for both Germany and the UK, are statistically significant at $\alpha = 0.01$.

VAR vs VECM In this section SSA is compared with classical time series models; in this case using the VAR and VECM models. Following the analysis of (Gupta & Kabundi, 2011) VAR models were selected. They were all found to have a regressor lag 1. All models were tested for the presence of a constant term, a trend, and both together but only the presence of a trend was found to be significant. Moreover, a VECM (as a restricted version of a VAR model) is also selected, for comparison, to account for cointegration in the time series. The RMSEs obtained by these models are somewhat larger than those estimated by the SSA based models. Specifically, the VECM models outperform the VAR models for France and the UK (but the differences are not found

⁵ Different cointegration models are applied and VECM was found to be the most accurate model for forecasting IPI time series.

⁶There are 28 different combinations for each country, the companion which yields the lowest in-sample RMSE is chosen.

to be significant⁷). On the other hand, for Germany, the VAR models outperform the VECM models.

MSSA-8 vs MBSSA-8 Here the case where a multivariate time series forecast is required of all eight time series is examined and so the models MBSSA-8 and MSSA-8 are compared. Referring to Tables 4.11-4.13, MBSSA-8 considerably outperforms MSSA-8, typically in the order of 1% to 2%. for horizons of up to 6 months. However, for horizons of a year ($h = 12$) the results are mixed with no model clearly outperforming the other in all situations.

Comparing MBSSA-2 with MBSSA-8 one can see that in most cases MBSSA-8 is superior which would seem to suggest that considering the 8 time series together would be advantageous.

ICA-MSSA-8 vs ICA-MBSSA-8 Next it is examined whether reducing the cointegration between the series prior to modelling, using ICA, leads to significant improvement. Indeed, significant gains in forecasting the industrial prediction are achieved. Overall, (Tables 4.11-4.13) ICA-MBSSA-8 shows 0.30, 0.40 and 1.6 reduction in comparison with MBSSA-8 for one-step-ahead forecasts, and 1, 2.9 and 4.7 reductions for $h = 12$. For medium runs, $h = 3$ and 6 the results shows an average reduction of 1 and 1.4 in RMSE for three countries. Last but not least, the GN test confirms the results obtained by the combined model are statistically significant at $\alpha = 0.01\%$ level.

Empirical cumulative distribution function Figure 4.7 shows the empirical c.d.f. for the absolute errors. The c.d.f. of the absolute values of the out-of-sample errors (for all twenty-four series) are computed from the MSSA-2 and MBSSA-2, VAR, VECM, MSSA-8, MBSSA-8, ICA-MSSA-8 and ICA-MBSSA-8 forecasts. Most noticeably of all, it can be seen that for $h = 1, 3, 6$ and 12, the ICA-MBSSA-8 forecasting errors are stochastically smaller than those of the VAR and VECM models. In addition, the forecast errors estimated by Bayesian MSSA models are shown to be much smaller than their MSSA counterparts.

4.4.3 Summary and conclusion

In this chapter, we developed a Bayesian model for the linear recurrent formula of the embedding stage in singular spectrum analysis. We proved theoretically that when a structural break happens to a time series the LRF, as an underlying assumption of SSA, does not coincide with a recurrent continuation of the series pre-break. Accordingly, in

⁷Granger Newbold test

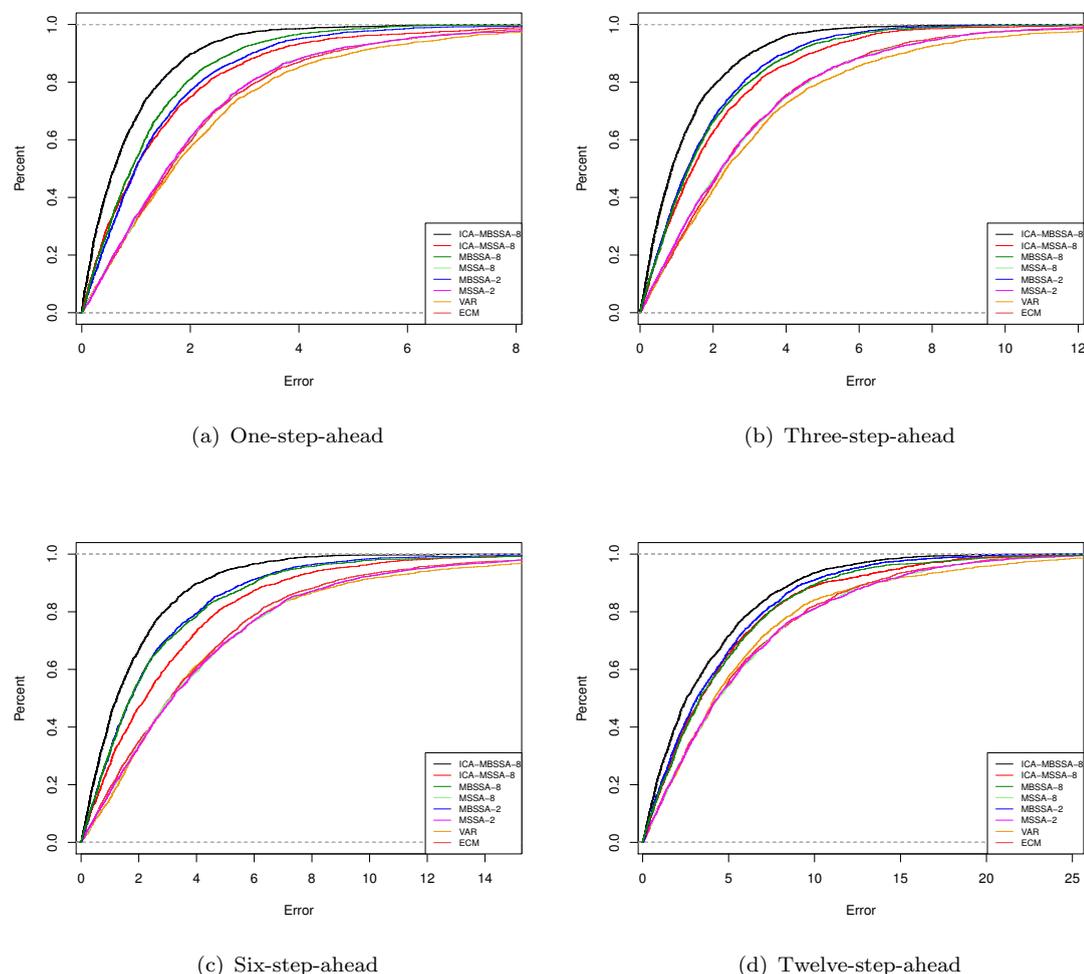


FIGURE 4.7: Empirical cumulative distribution functions of absolute values of forecast errors for ICA-GSSA-8, ICA-MSSA-8, MBSSA-8, MSSA-8, MBSSA-2, MSSA-2, VAR and VECM.

our model, called MBSSA, we propagate coefficients as a function of the state vector to involve their dependent movement in the LRF. The propagation coefficients can then mathematically explain their time evolution as well as their state position. We believe our model to be superior to the SSA technique in terms of dealing with an exponential decay in the LRF post shock.

The performance of the proposed model is assessed using both synthetic and real data (industrial production series) including a structural break. Of the four methods compared (MBSSA, MSSA, VAR and VECM) MBSSA was the most accurate method for forecasting horizons of up to a year. We were also interested to see how the model would perform in the presence of cointegration between all series (eight series for each country), given the promising results of the bivariate model. Our results showed that the model performed poorly compared to its bivariate counterpart. An obstacle to the

performance of this model is a lack of strong separability in SSA. Therefore, we applied ICA-MBSSA and found a large increase in performance.

Chapter 5

Latent and Reduced Space Multivariate Singular Spectrum Analysis

5.1 Introduction

In Chapter 4, a dynamic LRF was proposed based on a state dependent model. It was concluded that the state parameters allow the LRF parameters to recursively evolve based on the observations, and this can improve the forecasting accuracy significantly. The core element of the SSA algorithm are the eigenvectors of the trajectory covariance matrices as these determine the coefficients of the LRF. Selecting the correct number of eigenvectors will also provide a suitable series decomposition. However, when multiple time series are accompanied with common modes, the optimality of the SVD does not help to separate these common components due to a lack of strong separability (Golyandina et al., 2001) (Section 2.1.1.2). Therefore, special rotations can be found to satisfy some additional optimality criterion at the decomposition step (Golyandina & Zhigljavsky, 2013). The analysis in Chapter 4 indicates that reducing the cointegration between the series prior to modelling, using ICA, leads to significant improvement in forecast accuracy. However, ICA is a much less stable procedure than SSA therefore it is recommended by (Golyandina et al., 2001) that it should not replace the SVD entirely. On the other hand, in Chapter 4 we observed that bivariate forecasts (MBSSA-2 and MSSA-2) significantly outperform multivariate forecasts (MBSSA-8 and MSSA-8). The common argument above is that a combination of time series (either bivariate or by transform via ICA) may improve forecast accuracy and alternately may degrade it significantly. Therefore, a natural question arises which combination of time series support each other and which dont? In this Chapter we investigate this question using the USA unemployment rate series. This series is ideal for this task as the (USA) states

are approximately uniformly geographically distributed with neighbouring states often having common characteristics but also with anomalies (for example California, Florida and New York being prime economic states with agricultural states in between).

Unemployment is among a number of critical variables whose evolution has been continuously subject to close analysis by economic authorities and academics alike. Indeed, there is a large body of literature that deals with the estimation of (econometrics and time series) models aimed at understanding the determinants of this variable, and evaluating their ability to produce accurate forecasts; a few recent examples include, inter alia, (Milas & Rothman, 2008), (Lahiani & Scaillet, 2009), (Fendel et al., 2011), (Trendle, 2002), (Schanne et al., 2010) and (Ball et al., 2015).

Location as a determinant of economic growth plays an important role in finding spatial interdependencies between states, (Elias & Rey, 2007). On the other hand, spatial dependency as a determinant of movement and transactions determines an association between the commuting flow variables and spatial configuration variables (Anselin, 2013). Therefore, the essential purpose behind studying spatial econometrics time series, such as unemployment rates, is not only answering which series support each other but also their geographical homogeneity (or lack of). For example, New York and California tend to have same level of economic activity while they have by far the largest distance.

5.2 Spatial weights

In this Section, it is hypothesised that a spatially weighted series can improve the forecasting accuracy of each single time series of unemployment rate. To test the hypothesis, a bivariate time series, $(Y_N^{(1)}, Y_N^{(2)})$ is considered as a combination of the original series (i.e. the unemployment rate of the dependent state) and a spatially weighted series, respectively. The spatially weighted series includes regional information using geographical characteristics, such as distance and boundaries. This allows the model to incorporate the spatial dependency and regional heterogeneity into the forecasting procedures.

To construct the additional explanatory variable as a spatially weighted series, we first measure an individual weight for each region (as a spatial weights matrix) and then obtain a weighted average by assigning the individual weight to its related time series. There are different ways to determine the spatial weights matrix. In this study we examine the inverse distance and combined distance-boundary weights.

Inverse Distance Weights

It is apparent that the shorter the distance between two regions the more intense their connection. Specifically, the Inverse distance weight is considered one means of incorporating spatial information (Awichi & Müller, 2013) as:

$$w_{i,k} = \frac{1/d_{i,k}}{\sum_{i=1}^n (1/d_{i,k})}, \quad (5.1)$$

where $d_{i,k}$ is the distance from region i to the target location k . Therefore, the spatially weighted series is equivalent to $Y_N^{(2)} = \sum_{i=1}^N w_{i,k} Y_{N,i}^{(1)}$.

Combined Distance-Boundary Weights

Another geographical characteristic between regions is their boundary length (Cliff & Ord, 1969). The boundaries shared between neighbours can have a significant impact on recognition of spatial influence. The original study on spatial autocorrelation by (Cliff & Ord, 1969) proposed the best weighting scheme as the combination of power-distance and shared-boundary:

$$w_{i,k} = \frac{l_{i,k} d_{i,k}^\alpha}{\sum_{k \neq i} l_{i,k} d_{i,k}^\alpha}, \quad (5.2)$$

where $l_{i,k}$ denotes the length of the shared boundary between state i and k . Any positive integer can be chosen for α , typically $\alpha = 1$ or $\alpha = 2$, the larger α , the greater the weight on the distance (Cliff & Ord, 1969). After incorporating spatial information, we apply MSSA and MBSSA on the original time series and this additional explanatory variable, $(Y_N^{(1)}, Y_N^{(2)})$, for each (USA) state. To assess the impact of spatial weights on forecasting unemployment rate series we tested our model, MBSSA with two spatial versions of MSSA, as well as univariate SSA.

5.2.1 Data, descriptive statistics and graphs

In this section, we give a brief summary of unemployment data. The data used in this study is the seasonally adjusted monthly unemployment rates for 48 states of the USA between January 1976 to December 2013 ($N = 456$). The source of the data is the Federal Reserve Economic Database (FRED) provided by the Federal Reserve Bank of St. Louis.

Figure B.1 (see Appendix B) gives us a clear overview of the time series structure. For example, periods of business cycles, including both expansion and contraction are identifiable. As can be seen, most of the states show a short period of growth in late 1982 and beginning of 1983. Moreover, between the third quarter of 1991 and the second quarter of 1992, the rate of unemployment has marginally increased, for instance in Florida, Connecticut and Massachusetts.

Despite some minor fluctuations, for almost all the states, the last significant growth was between 2009 and 2010. It should also be noted that there are strong local variations.

Unemployment rates by State, seasonally adjusted, December 2013

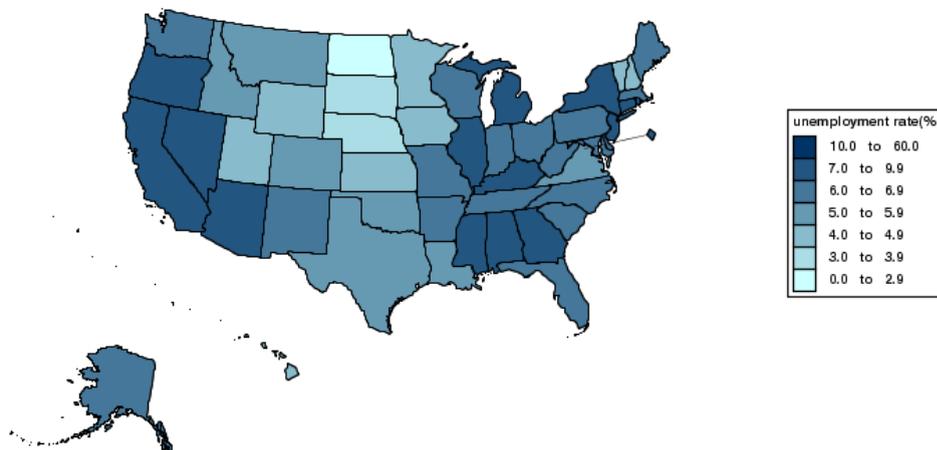


FIGURE 5.1: Unemployment rates for the United States reported by the USA Bureau of Labour Statistics.³

For instance, in the Northeast, Alabama, Tennessee and Kentucky share a common structure over the entire time span. The rate of unemployment also follows a similar pattern in “West Virginia, Ohio and Pennsylvania” and “Idaho and Utah” in the West of the United States.

A choropleth map, Figure 5.1, is an alternative way to examine the impact of various geographical locations on the series in Unites States. It shows the average of seasonally adjusted unemployment rates for December 2013 based on the Bureau of Labour statistics report ¹. According to the map, the Northeast USA has the highest unemployment rate, with 7 states out of 9 states having an unemployment rate above 6%. 13 states out of 17 states in the South also have an unemployment rate above 6%. The rate of unemployment is equally distributed among the states in the West. Alternately, Midwest states shows 6 out of 11 states with the lowest unemployment rate on average.

In fact, the above report can be confirmed by Table B.1 in Appendix B. It provides summary statistics for the regional unemployment rates of the USA. These show that the average unemployment rates range from 3.8% in South Dakota to 8.3% in West Virginia over the thirty seven year period. Moreover, the sample standard deviation (SD) indicates 0.8 volatility in South Dakota and 3.1 in West Virginia, as the lowest and highest volatility among all the states, respectively. A box-plot of the data, Figure 5.2, is a more convenient way of graphically depicting minimum and maximum, mean, inter-quartile and also variability outside the upper and lower quartiles of the series. Most of the series display a significant number of outliers at high maxima or low minima, like Alabama, Louisiana, Mississippi, North Carolina, Ohio and South Carolina. Despite

¹<http://data.bls.gov/map/MapToolServlet>

statistics are positive. Consequently, we can conclude that there is a positive spatial autocorrelation between those states.

The local Moran's Index allows for decomposition of the global indicators into the contribution of each individual observation (Anselin, 1995). It means they provide a local instability in overall spatial association of regions. A local Moran statistic, I_i , can be calculated as:

$$I_i = \frac{n(Y_i - \bar{Y}) \sum_j w_{i,j}(Y_j - \bar{Y})}{\sum_j (Y_i - \bar{Y})^2}. \quad (5.4)$$

The local Moran statistic measures the local deviations from the global pattern of spatial association (Scrucca, 2005). It is also shown in (Moran, 1950) that the average of I_i is equivalent to the global I , up to a factor of proportionality.

The Moran scatterplot of the US unemployment rate, for both inverse-distance weights (ID) and Border-Distance weights (BD), is given in Figure 5.3. The four quadrants are associated with four types of spatial association. As can be seen the majority of significant positive associations are in the upper-right quadrant especially for Border-Distance weights. It means districts with high unemployment rates are connected to those with high values. On the other hand, states with a lower rate of unemployment are classified into the lower-left quadrant. The obtained value of I can be used to test the null hypothesis of no spatial autocorrelation between districts. Both spatial weights, ID and BD, show a significant spatial autocorrelation amongst the whole area by rejecting the null hypothesis. The greater value of the I -statistics for BD ($I = 0.617$) also confirms that there is a stronger spatial association when weights are related to both border and distance.

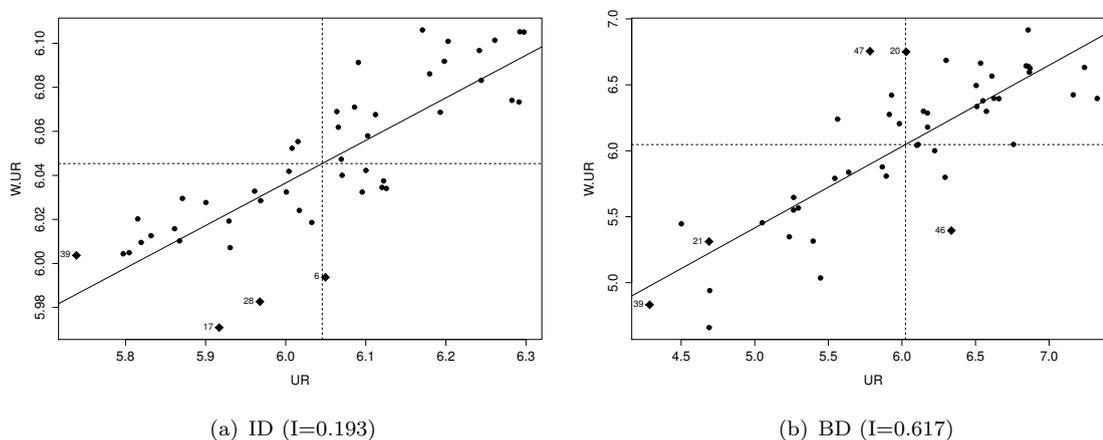


FIGURE 5.3: Moran Scatterplot for United States Unemployment Rate.

The results of the local Moran test are presented in Table B.2 (see Appendix B). Among 48 states, 20 states show statistically significant spatial interdependency in the case of an

Inverse-Distance weight. Moreover, over 25 states show a strong spatial association at 5% significance level in the case of a Border-Distance weight. These findings essentially provide support for the key argument of our study, which is that using spatial weights for forecasting unemployment rates based on the multivariate model may be advantageous.

5.2.1.1 Empirical Results

In this section, we compare the performance of out-of-sample forecasting using SSA, BSSA, MSSA and MBSSA for bivariate series with Border-Distance weights as well as Inverse-Distance weights. All comparisons are made in terms of prediction RMSE with respect to MSSA, and univariate SSA as the benchmarks. In addition, comparisons are made at four different horizons: monthly, quarterly, half-yearly, and yearly. The last third part of each series is reserved as an out-of-sample data.

Tables B.3, B.4, B.5, B.6, B.7, and B.8 in Appendix B display detailed information on the RMSE and RRMSE for the unemployment rate of the 48 states of the USA. Table 5.1 presents the summary statistics of the RMSE, RRMSE and number of significant forecasts obtained by each method across all horizons. There is no clear indication whether BD provides better forecasts than ID. Although, it seems BD is slightly better than ID. Meanwhile, the summary shows that BSSA is significantly superior to SSA for both spatial models, ID and BD. Comparing results of MSSA with univariate SSA, however, indicates a slight improvement in forecast accuracy. According to table 5.1, MBSSA provides the most accurate forecasts, showing a considerable reduction in RRMSE. In both cases, BD and ID, MBSSA results in an average gain of 32%, 44%, 45% and 38% over MSSA for $h = 1, 3, 6, 12$ which would seem to suggest empirically that a Bayesian multivariate model is superior in this case. Comparing results of MBSSA with univariate BSSA, indicates an average reduction of 32%, 13%, 12% and 16% for each horizon. We also report the statistical significance of the RMSE (relative to each comparison), and these show that 94% of MBSSA forecasts for both BD and ID, are statistically significant at $\alpha = 0.01$. In the medium and long term, 96% of MBSSA forecasts show a statistically significant result. The improved forecast accuracy of MBSSA over BSSA suggests that unemployment time series can support each other. The natural question then arises whether all the supporting information is contained in the geographical distribution of states or other supporting factors exist as will now be explored.

TABLE 5.1: Summary statistics for out-of-sample forecasting accuracy measures for unemployment rate series.

Measures	Methods	1	3	6	12
RMSE	SSA	0.12	0.33	0.62	1.12
	BSSA	0.11	0.24	0.40	0.86
	ID-MSSA	0.12	0.32	0.63	1.07
	BD-MSSA	0.12	0.32	0.60	1.02
	ID-MBSSA	0.07	0.18	0.34	0.64
	BD-MBSSA	0.07	0.17	0.33	0.64
RRMSE	ID-MSSA/SSA	0.98	0.99	1.03	0.96
	ID-MBSSA/BSSA	0.68	0.87	0.88	0.84
	ID-MBSSA/ID-MSSA	0.69	0.55	0.53	0.60
	BD-MBSSA/BD-MSSA	0.67	0.58	0.58	0.64
Sig at 0.01	ID-MSSA/SSA	0.00	2.00	2.00	7.00
	ID-MBSSA/BSSA	48.00	41.00	40.00	44.00
	ID-MBSSA/ID-MSSA	45.00	46.00	48.00	48.00
	BD-MBSSA/BD-MSSA	47.00	48.00	48.00	46.00

5.3 Projection pursuit using Joint Diagonalisation

As mentioned in chapter 2, there are two forms of the block trajectory matrix in MSSA along with two forecasting procedures, Horizontal and Vertical. A horizontal stack of Hankel matrices enables us to have various K_i and different series length N_i , however similar L_i for all series, while a vertical stack of Hankel matrices, enables us to have various window lengths L_i and different series lengths N_i , however similar K_i for all series.

(Hassani & Mahmoudvand, 2013) compares the structures and limitations of horizontal and vertical MSSA forecasting algorithms from different perspectives such as the series length, the value of window length (L_i), the number of non-zero singular values obtained from the block trajectory matrix, and the coefficients of the LRF. (Hassani & Mahmoudvand, 2013) also discuss matched and unmatched components between multivariate time series. They show these components sometimes support each other and sometimes don't. Therefore, they attempt to examine the block diagonal matrix of multivariate time series which does not restrict the dimension of covariance matrices, C_i . However, this form of covariance matrix is equivalent to M univariate cases which results in the same eigen-decompositions as univariate SSA.

At this point it is instructive to define exactly what we mean by *matched* and *unmatched* components (as these terms are not explicitly defined in (Hassani & Mahmoudvand, 2013)). As we will be comparing the models based on their forecast accuracies given various combinations of series, in one sense we mean the Granger Causality between series (see (Hamilton, 1994) Chapter 11 for an in depth discussion with respect to multivariate analysis.); i.e. that inclusion of one or more time series in a model may improve

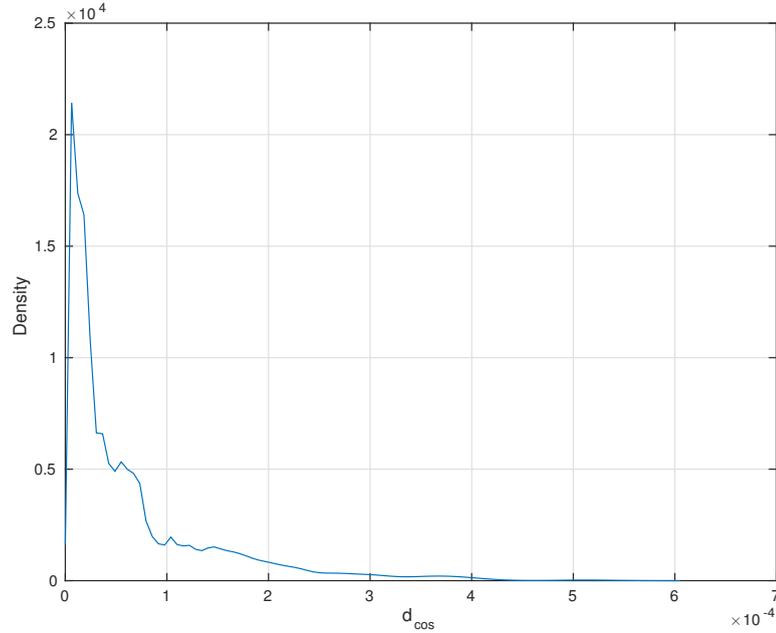


FIGURE 5.4: Distribution of cosine distances between the strongest eigenvector of every state

a forecast of the dependent time series. With specific reference to SSA based models, the LRF coefficients are determined once the eigenvectors of the covariance matrix have been determined. If two time series have two (true) eigenvectors which are similar then it is possible that both samples combined would result in a better estimate of the true eigenvector of the dependent time series. In such a case we say these time series have a matched component. The difficulty lies in the definition of *similar* and we cannot define, for example some threshold angle between vectors, below which the time series reinforce each other. On the contrary, we have found that the unemployment time series all share very similar eigenvectors but that small angles between these vectors can lead to significant forecast differences as will be shown. Figure 5.4 shows the distribution of angles between all the eigenvectors of the 48 time series for the eigenvector $u \approx [11\dots 1]^T$.^{4,5} As can be seen the angle between them is small with a maximum cosine distance of 5.2×10^{-4} .

Suppose we have a set of covariance matrices for which we wish to find a set of orthogonal matrices U such that the coefficients of projection onto this new basis are as diagonally dominant as possible. The problem can be solved efficiently by using a simultaneous decomposition method which looks for the communality among the matrices. One way of doing that is to consider a common set of eigenvectors, between all covariance matrices of multivariate time series i.e. via Joint Diagonalisation (See Section 2.1.1.2).

⁴The cosine distance is used $d_{cos} = 1 - \frac{u_a u_b}{\sqrt{(u_a u_a^T)(u_b u_b^T)}}$.

⁵Normalised to one. Note this is the eigenvector which holds the trend component and is the strongest in each series.

In this section, as an alternative to the SVD, the second step of SSA, is replaced by JD. The key differences between SVD in MSSA and JD, is that the eigenvector in the former is a summation of covariance matrices $U_{\sum_i X_i X_i^t}$ while the latter targets a common vector as an average eigenvector of covariance matrices $\bar{U}_{\{X_i X_i^t\}}$. In fact JD can directly search for the common eigenspace. It is worth noting that, JD MSSA is essentially a variant of Horizontal MSSA. In general the eigenvectors of a sum of matrices can differ significantly from their average eigenvectors (via JD), although when the covariance matrices of multivariate time series are quite similar then the eigenvectors of both approaches are similar.⁶

From Section 2.1.1.2 the JD of a set of matrices seeks to minimise:

$$\delta_i = \text{off}_2(Q_i) = \sum_{k \neq j} Q_i^{k,j^2} \quad (5.5)$$

where off_2 is defined as the sum of the off diagonal elements and called the deviation from C_i , $Q_i^{k,j}$ is the k^{th} row and j^{th} column of Q_i . Given the average eigenstructure of the covariance matrices an average covariance matrix may be constructed from the eigenvector decomposition as:

$$\bar{C} = \bar{U} \bar{Q} \bar{U}^T \quad (5.6)$$

Where \bar{C} is a matrix in which the entries represent the average weight at the covariance matrices of each time series and \bar{Q} is the average of diagonals of C_i projected onto \bar{U} which might also be called the average eigenvalues (Fay & Yoneki, 2011). In addition, the statistical distribution of deviations, δ_i , from the average provides us with an interesting picture about the underlying structure of the covariance matrices as will be shown in the following section.

5.3.1 Empirical Results

In this Section we compare the spatial SSA models with SSA based on JD in order to determine if this non-spatial clustering of the time series is advantageous. The results are compared with those obtained via MSSA and MBSSA (all 48 states together) in terms of their accuracy at predicting short, medium and long-term horizons. The second part looks at the distribution of δ_i and presents the first of 5 alternate clustering maps of the USA.

5.3.1.1 JD versus SVD

In this section JD-MBSSA and MBSSA are compared given that both use just a single eigenspace (the JD average eigenvectors and the eigenvectors of the sum of the 48

⁶The eigenvectors of $A + A$ equal that of A equal that of $\text{JD}(A, A)$.

covariance matrices respectively). Thus, both JD-MBSSA and MBSSA consider all 48 series together. Tables B.15 to B.20 compare the out-of-sample RMSE and the RRMSE results for MSSA, MBSSA, JD-MSSA and JD-MBSSA.

Table 5.2 presents the summary statistics of the RMSE and RRMSE at each horizon. The result indicates JD-MBSSA does not perform well compared to its counterpart, MBSSA (and JD-MSSA compared to MSSA). The average RMSE's for four horizons are 1.40, 1.63, 1.53 and 1.43.

It might be concluded that SVD performs better than JD for this particular data sets. However, JD provides a stacked matrix with the diagonal matrices (or equivalently an array with the diagonal matrices) as the same format as the covariance matrices. The stacked matrix can be then used for grouping covariance matrices into clusters based on their common structure to understand the processes underlying them. As discussed earlier, this common structure can be found by looking at the distribution of δ_i .

TABLE 5.2: Summary statistics (JD, Tensor and SOEM clusters vs SVD).

Steps	RMSE								RRMSE		
	MS	MBS	JD-MS	JD-MBS	Ten-MS	Ten-MBS	SOEM-MS	SOEM-MBS	$\frac{JD-MBSSA}{MBS}$	$\frac{Ten-MSSA}{MBS}$	$\frac{SOEM-MBS}{MBS}$
1	0.12	0.08	0.12	0.12	0.13	0.08	0.10	0.07	1.40	1.01	0.88
3	0.37	0.20	0.38	0.32	0.36	0.20	0.27	0.18	1.63	1.01	0.87
6	0.77	0.40	0.83	0.60	0.84	0.40	0.58	0.37	1.53	1.01	0.92
12	1.68	0.76	2.09	1.08	2.76	0.76	1.22	0.73	1.43	1.00	0.96

Note: The following abbreviations are used to take up less space: MS(MSSA), MBS(MBSSA) JD-MS(JD based MSSA) and Ten-MS(Tensor based MSSA) and so on.

5.3.1.2 Clustering states on δ_i

The purpose of cluster analysis of δ_i is to determine the inner structure of clustered states which can be an alternative procedure to partition multivariate time series to have better forecast performance. The cluster analysis essentially shows geography of USA employment. Here the most commonly used methods of clustering, such as Gaussian mixture model (Dempster et al., 1977; Day, 1969), kmeans (Lloyd, 1982; MacQueen et al., 1967) and Hierarchical clustering (Johnson, 1967; Lance & Williams, 1967) are used to discover subclasses of USA states. Note that the optimal clustering is subjective and depends on the method used for measuring similarities and dissimilarities and related parameters used for partitioning.

Gaussian mixture model The analysis starts with estimating the underlying probability distribution of off diagonal elements squared via a Gaussian Mixture Model (GMM) (McLachlan & Peel, 2004) in which the problem of choosing the number of modes (components) can be reformulated as a statistical model choice problem. In

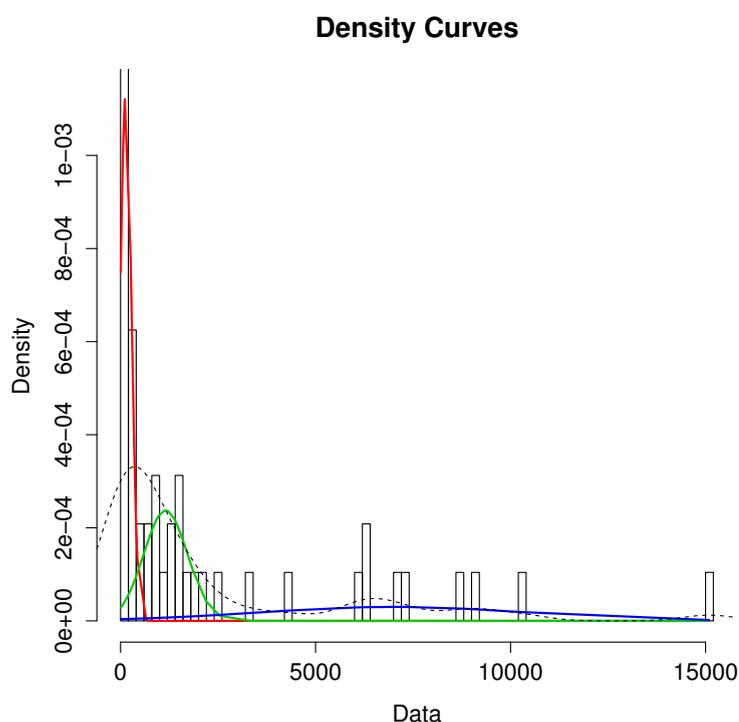


FIGURE 5.5: Distribution of δ_i (kernel smoothing is employed for the overall average.)

this context, δ_i are viewed as coming from a mixture of probability distributions, each representing a different cluster.

Examining the distribution of deviations, δ_i , from the average, a more interesting behaviour may be observed. Figure 5.5 illustrates the empirical distribution of $\delta_i, i = 1, \dots, 48$. It is evident that the distribution is multi-modal; i.e. the underlying process has different modes of operation. It also divides δ_i in a way that conforms reasonably well with intuitive ideas of natural clustering. Figure 5.5 shows a mixture of 3 different modes⁷. This is particularly useful as it allows the multivariate system to be characterised by different modes of behaviour at different times.

A cluster mapping of USA is also shown in Figure 5.6 which provides sectoral and cross-sectoral statistical analysis of regional data. This map creates a dataset on the presence of clusters across geographical characteristics according to the clusters provided by GMM. According to the map, there are 3 (blue, red and yellow) clusters concentrated in a subset of geographic areas however these clusters are scattered and do not follow a particular pattern. For example, states like New York, Florida and California tend to have same types of economic activity while having by far the largest distance. Conversely, there is clearly a geographic contiguity in the clusters which suggests neighbouring states economies are linked.

⁷ The model with the lowest Bayesian information criterion (BIC) is preferred



FIGURE 5.6: States clustered with GMM.

Kmeans clustering Another way to cluster states is via Kmeans (Sammut & Webb, 2011). It basically measures dissimilarity based upon Euclidean distance. The Kmeans algorithm requires one to specify the number of clusters, K . However, there are some specific measures that can be used to examine the number of clusters in Kmeans (Sammut & Webb, 2011). Here, we set $K = 3$ so a fair comparison may be made with the other techniques in this Chapter. Figure 5.7 maps the regional clusters provided by Kmeans.



FIGURE 5.7: States clustered with Kmeans.

According to Figure 5.7, 44% of states in West and Midwest are in one cluster, compared to 22% in South (and Washington and Wyoming from West). On the other hand, New York, Florida, Nevada and Oregon belong to the same cluster which is very close to those provide by GMM. Again we see a large degree of contiguity between the states but there are also some states clustered together which would not appear to be a natural cluster. For example Vermont is grouped with the southern states and Florida would



FIGURE 5.9: States Hierarchically clustered.

5.3.1.3 Forecasting clustered states

Using the above we showed there exist a definite spatial correlation among states using unemployment data. Although, there is a large variation in the clusters formed by the above mentioned clustering algorithms. All seems to confirm the level of economic activity and neighbouring between states is relative to their unemployment rate.

The detailed results forecasting USA unemployment rates based on the clusters provided by GMM, Kmeans and hierarchical is reported in B.9, B.10, B.11, B.12, B.13, and B.14 in Appendix B. MSSA and MBSSA using clusters are compared with MSSA and MBSSA using all 48 states together for each clustering. A summary of those results is provided in Table 5.3. There is not a clear indication of which grouping procedure is most likely to provide better forecasts. Meanwhile, the summary shows that MBSSA is still significantly superior to MSSA. The results of comparing km-MBSSA and hierarchical-MBSSA with MBSSA, however, reveal using clustered states could not improve forecasting accuracy of MBSSA. The average RRMSE shows GMM is slightly better than Kmeans and hierarchical. Overall, it might be concluded that for all horizons, difference between these three techniques is negligible.

TABLE 5.3: Summary statistics for clustering based on JD.

Steps	RMSE					RRMSE				
	MBS	km-MBS	km-MS	GMM-MBS	GMM-MS	Hier-MBS	Hier-MS	$\frac{km-MBS}{MBS}$	$\frac{GMM-MBS}{MBS}$	$\frac{Hier-MBS}{MBS}$
1	0.08	0.08	0.12	0.08	0.12	0.08	0.12	1.01	0.98	1.02
3	0.20	0.20	0.38	0.19	0.37	0.20	0.37	1.01	0.95	1.01
6	0.40	0.40	0.85	0.38	0.79	0.40	0.77	1.01	0.96	1.01
12	0.76	0.76	2.65	0.75	1.83	0.77	1.69	1.00	0.99	1.01

5.4 Projection pursuit with Tensor decomposition

As discussed before, MSSA looks for orthogonality between time series to decompose covariance matrices together. Although it does not expose communality between their modes. Therefore, we used JD to find a common set of eigenvectors, between all covariance matrices of multivariate series. Analysis provided in the previous section showed even a common set of eigenvectors can not solve the problem since those common modes between different groups of time series are still supporting each other.

The problem with JD can be expressed in two ways. First, JD can only allow us to identify common modes between all covariance matrices and second using δ_i as a feature (scalar/vector) is not an ideal choice as it can only provide one dimensional clustering. To be more specific, the distribution of off diagonal squared, δ_i , only relies on how far we are from the average eigenspace and therefore different clusters which have the same distance from that average are indistinguishable. In this part, we bring a multidimensional view to decompose the covariance matrices using a more generalised format of JD, called tensor analysis.

5.4.1 Tensor Analysis

(Lathauwer, 2011) shows that in the third-order case the computation of the PARAFAC/-CANDECOMP of a tensor \mathcal{X} is equivalent to the simultaneous diagonalisation of its matrix slices. In this part, we aim to capture the latent structure underlying the data via a Tucker transform. The covariance matrices of multivariate series are considered as tensors instead of more conventional matrix representations. In this approach, the multi-way data is decomposed using the Tucker decomposition without constraints to obtain basis factors and consequential features from the core tensors.

Figure 5.10, diagrammatically shows the 3-way decomposition of the M covariance matrices $C_{i,j}^{(m)}$ as performed via the Tucker transform. The input tensor is a $L \times L \times M$ tensor with three modes and this is then decomposed into three matrices and a core tensor. The core tensor is a lower dimensional approximation of the original tensor in which the projection amplitudes of the 3-modes are given by $U^{(1)}$, $U^{(2)}$ and $U^{(3)}$.

If we examine the core tensor sliced along the 3^{rd} mode then the first slice represents the strongest direction in the slices of the original tensor. Lets assume that there are three clusters in the data (in the sense that within each cluster there is a common eigenspace) and that the *blue* cluster is the dominant cluster (Figure 5.10). Then the first slice of the core will contain approximately the (average) eigenvalues of the first cluster and $U^{(1)}$ and $U^{(2)}$ will contain their eigenvectors. The second slice of the core will contain the coefficients of projection of the second cluster onto the eigenspace of the first cluster. As

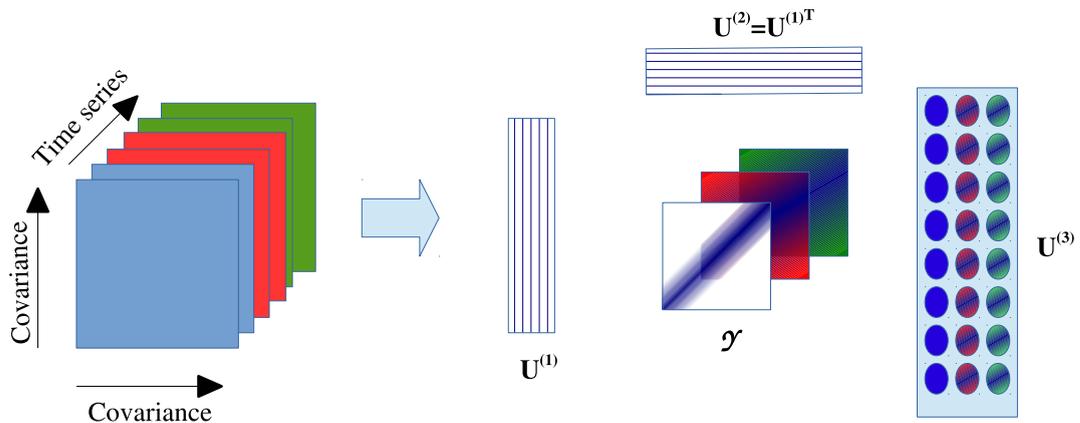


FIGURE 5.10: Tucker transform.

this is not the eigenspace of the second cluster these coefficients will contain strong off-diagonal elements. Similarly, the projections of the third cluster onto the eigenspace of the first will not be diagonal and is contained in the third slice. Thus, in Figure 5.10 the core tensor is shown as having a diagonally dominant first slice and the colours mixing in the subsequent slices. $U^{(3)}$ gives the projection of each time series into the slices. In an ideal situation where the clusters are distinct $U^{(3)}$ would be block diagonal containing just 1 to indicate which core represents which cluster. However, as seen in Figure 5.4, for the unemployment rate times series used there is a large similarity between the time series. Thus the first slice of the tensor contains the average directions, and the second and subsequent slices contain projections of the deviations from the average of each time series. $U^{(3)}$ thus contains the extent to which each time series requires these deviations to produce a reasonable reconstruction and so $U^{(3)}$ may be considered as a projection of the time series into a 3D space/co-ordinate system which separates the time series and so may be used for clustering.

5.4.2 Empirical results

In this section we replace SVD with Tucker transform, on a tensor-sized $L \times L \times 48$ covariance matrices. Different values for the window length, L , are examined and we found $L = 3$ is the suitable choice for all 48 states combined. Our analysis also confirms the above discussion that the first core carries most of the power.⁸ The second core shows that the sum of off-diagonal elements is indeed stronger than the diagonal. On the other hand, the primary eigenvectors of the two projections $U^{(1)}$ and $U^{(2)}$ are similar.

⁸The sum of diagonal of each core is its power.

TABLE 5.4: Summary statistics comparing spatial SSA and tensor SSA using a single eigenspace/cluster.

Steps	RMSE				RRMSE	
	Tensor-MSSA	Tensor-MBSSA	BD-MSSSA	BD-MBSSA	$\frac{\text{Tensor-MBSSA}}{\text{BD-MBSSA}}$	$\frac{\text{Tensor-MSSA}}{\text{MSSA}}$
1	0.13	0.07	0.12	0.07	0.96	1.07
3	0.36	0.18	0.32	0.17	1.00	1.11
6	0.84	0.37	0.60	0.33	1.10	1.37
12	2.76	0.76	1.02	0.64	1.18	1.45

Table 5.4 only compares the results with those obtained by BD in Section 5.2. The difference between the tensor based approach and the BD approach is marginal with the border distance giving marginally superior results.

Additionally, Table 5.2 shows the differences between of clustered series via $U^{(3)}$ and MBSSA is negligible.

5.5 Projection pursuit with a self-organising map

In Section 5.3 we noted that two time series can be considered to have matched components if their eigenvectors are *similar*. As seen in Section 5.4 the Tucker transform based approach is not ideal essentially because we have only one option for selection of $U^{(1)}$ and so can never produce a set of core slices that are all diagonally dominant. Due to the small angles between the eigenvectors of each cluster we require a technique that can produce clusters with closely aligned bases. Rather what we seek is a set of clusters whose centres are defined by orthonormal eigenspaces which are non-orthogonal to each other (i.e. each cluster may contain very similar eigenvectors and in some cases we expect the angles between them to be quite small).

5.5.1 Self Organising Eigenspace Map

Specifically, K clusters are sought, $S_{1:K}$, each defined by a set of orthonormal vectors, $\{U_k\}_{k=1:K}$ such that the off-diagonal coefficients of the covariance matrices, C_i , projected onto the orthonormal set within each cluster is minimised as:

$$\operatorname{argmin}_{\{U_k\}_{k=1:K}} \sum_{k=1}^K \sum_{\forall C_i \in S_k} \langle U_k, C_i \rangle \quad (5.7)$$

where $\langle U_k, C_i \rangle$ denotes some distance measure between the eigenvectors of a cluster and the covariance matrix of an American state. The optimisation in Equation (5.7)

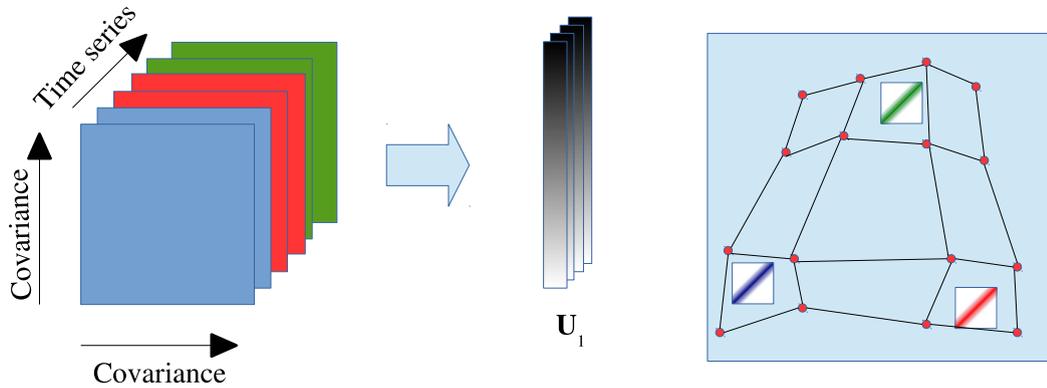


FIGURE 5.12: self-organising eigenspace map.

depends on the number of clusters, and in addition an optimal set of eigenvectors must be found; as neither is known it would be desirable to infer both during the solution.

Here, we employ the *Self Organising Eigenspace Map (SOEM)* algorithm (SOEM) (Fay & Rahmani, 2017) which is inspired by the Kohonen neural network but which incorporates matrix inputs and employs a weighted JD as the core update mechanism. Figure 5.12 shows diagrammatically the SOEM proposed for clustering the covariance matrices of the times series.

5.5.1.1 SOFM vs SOEM

A Kohonen neural network consists of a grid of topologically distributed nodes each with an associated function. An input is presented to this grid (i.e. the functions are evaluated for the input at every node) and the location of the maximum (or minimum) activation is called the *winner*. During training the winning nodes function (and crucially) its neighbours are updated so as to reinforce the activation. Assuming the network training converges, the distribution of the winning nodes reflects the topological ordering of the inputs, in addition, the nodes should place themselves such that the input which maximise them are sampled from the same underlying distribution as that of the input space (Tatoian & Hamel, 2016). In such a case the 2-D grid of nodes may be considered a lower dimensional mapping of the input space and as such useful for clustering. Specifically, the Kohonen map (also called the Self Organising Feature Map (SOFM)) and its variant, the SOEM consist of the following steps:

1. **The competitive step.** In an SOFM each node has an associated vector, $\vec{v}_{i,j}$ and given a vector input \vec{x} the winning node, s , is:

$$s = \underset{i,j}{\operatorname{argmin}} (|\vec{v}_{i,j} - \vec{x}|) \quad (5.8)$$

For an SOEM each node has an associated orthonormal basis, $\{U_{i,j}\}$, and the winning node, s , is that which aligns best with the projection space of the matrix input, C , as:

$$s = \underset{i,j}{\operatorname{argmin}} \operatorname{off}_2\{(U_{i,j}^T U_{i,j})^{-1} U_{i,j}^T C U_{i,j}\} \quad (5.9)$$

2. **The update step.** In an SOFM the node vectors are rotated towards the input vector according to:

$$\vec{v}_{i,j} \leftarrow \vec{v}_{i,j} - \nu(\vec{v}_{i,j} - \vec{x})h(s, i, j) \quad (5.10)$$

where ν is a gain term $\in [0, 1]$, $h(s, i, j)$ is a kernel which is (typically) monotonically decreasing with the distance from node i, j to the winning node s . For an SOEM the node bases are rotated towards the input matrices according to:

$$U_{i,j} \leftarrow \mathbf{J}_2\left(U_{i,j} | [\vec{v}_{i,j} U_{i,j} \quad h(s_1, i, j)C_1 \quad h(s_2, i, j)C_2 \dots h(s_M, i, j)C_M]\right) \quad (5.11)$$

where $\mathbf{J}_2(\cdot)$ is the joint diagonalisation of the old vector set with a scaled/weighted set of the input matrices and $h(s_k, i, j)$ is a monotonically decreasing function in the distance between $\{i, j\}$ and s_k . Note that for an SOFM each input presented updates $\vec{v}_{i,j}$ in an iterative fashion while for an SOEM all the inputs are used simultaneously to update a basis⁹.

3. **The iteration step.** After all inputs have been presented the algorithm typically changes and for example ν may be reduced (thus slowing the refinement of the node vectors to stop oscillations). Also, $h(s_k, i, j)$ may be changed, for example if $h(s_k, i, j) = \mathcal{N}(0, \sigma_h)$ is a Gaussian kernel then the width of this kernel may be reduced after each iteration (thus allowing a local rather global refinement of the solution). In the current research our SOEM employs a Gaussian kernel with $\sigma_h = 1/4$ width of the grid allowing each input to effect a large section of the grid.

Following step 3 the algorithm returns to step 1 and iterates until a stopping condition is found (in this research a set number of iterations is given although the algorithm may be stopped, for example, when the kernels effective radius falls below one). The SOEM is initialised by assigning random vectors to each node which are then adjusted to form a basis using Gram Schmidt orthogonalisation. In addition, each input is scaled such that it has unity norm. In the first iteration, the winning nodes are assigned randomly; this ensures the first iteration spans the entire eigenspace of the inputs.

⁹We experimented with iterative updates but found that the eigenspaces are adapted either too slowly or quickly to each input and so would not converge.

5.5.2 Empirical results

Figure 5.13(a) shows the first iteration demonstrating that the winning nodes are uniformly distributed. Figure 5.13(b) shows the second iteration and demonstrates that some clustering is already starting to appear. Figure 5.13(c) shows the final iteration; as can be seen there are 3 distinct clusters. The red cluster on the left the blue cluster and the yellow cluster in the middle which borders both larger clusters. Figures 5.14(a) to 5.14(d) show the values of $\text{off}_2(\|U_{i,j}, C\|)$ (i.e. the distances in Equation (5.7)) for indicative states. As can be seen California and New York have a lower value towards grid location (0,0) (the blue cluster) while Texas and Iowa prefer location (30,30). The key point in these figures is that the transition across the grid is smooth showing that the SOEM has achieved a topological ordering from the data.

In (Tatoian & Hamel, 2016) the authors propose two measures for evaluating the convergence of a Kohonen neural network. The first measure is the *topographic accuracy* defined as the average number of neighbours of winning nodes that came second:

$$\epsilon_{\mathcal{T}} = 1/n \sum_{i=1}^n 1/\|s^{(n)}, s_2^{(n)}\|_{\mathcal{H}} \quad (5.12)$$

where $\|s^{(n)}, s_2^{(n)}\|_{\mathcal{H}}$ is the distance between the location of the winning node s , for input n and that location which came second, s_2 . In addition, $\|\cdot\|_{\mathcal{H}}$, denotes the *Hamming* distance in the sense that it is 0 if s and s_2 are not neighbours. In a (ideal) topologically ordered space the winning and second place nodes would always be neighbours and this is the motivation for the measure in (5.12). The value of $\epsilon_{\mathcal{T}}$ for the map trained in this research is 0.1875 showing that approximately 20% of the time the 2nd best location is a neighbour of the winning node. However, this number does not concur with the general impression conveyed by Figures 5.14(a) to 5.14(d) because it does not take into account the noise present in an empirical ordering.

Instead we report the average distance to s_2 averaged over all the state which is 3.92. Indeed, the distance from the i^{th} place loser increases monotonically from the winner as shown in Figure 5.15. This demonstrates that a topological ordering has occurred as there is linear relationship between the place achieved and the distance to the winner.

5.5.2.1 Clustering with SOEM

Figure 5.16 presents the clustering of states given by the SOEM. It quite clearly shows a central band of red states and east and west coast states in a blue cluster. In addition, there is also a third yellow cluster of states located in the far South. From an intuitive point of view this clustering appears to make sense as states such as Florida, New York and California are clustered together while the more agricultural central states are also clustered together.

TABLE 5.5: Summary statistics (MSSA, MBSSA, and BD vs SOEM).

Steps	RMSE					
	MS	MBS	BD-MS	BD-MBS	SOEM-MS	SOEM-MBS
1	0.12	0.08	0.12	0.07	0.10	0.07
3	0.37	0.20	0.32	0.17	0.27	0.18
6	0.77	0.40	0.60	0.33	0.58	0.37
12	1.68	0.76	1.02	0.64	1.22	0.73

Note: The following abbreviations are used to take up less space: MS(MSSA), MBS(MBSSA) BD-MS(JD based MSSA) and SOEM-MS(SOEM based MSSA) and so on.

5.5.2.2 Forecasting clustered states

The overall forecasting results obtained by forecasting each cluster together is given in Table 5.2 in comparison with Table 5.1 which are combined in Table 5.5 below. The SOEM clusters lead to a lower RMSE than the MBSSA models in all cases and a similar performance to the BD-MBSSA models. It should be noted that SOEM BD-MSSSA comparison is not quite a fair comparison as the BD models essentially consist of pairs of time series individual to each state (i.e. equivalent to 48 clusters) while the SOEM is restricted to just 3 clusters. In addition the SOEM provides us with an excellent performance and *clustering* while neither the BD or the MBSSA models can provide clustering and the alternate clustering techniques cannot provide excellent performance.

5.5.3 Summary and conclusion

This Chapter concentrated on pooling time series together in such a way as to reinforce their common modes. As noted in Section 5.3 there is a large similarity/commonality between the time series modes in the data used here but yet pooling all time series together results in degradation of performance (Table 5.1). We investigated different means of pooling time series. The first was based on the physical location of the states themselves Section 5.2, the remaining Sections employed the data itself to extract some common pooling (Sections 5.3, 5.4, and 5.5).

The Tensor based, SOEM based and three JD based clusters shown in Figures 5.6, 5.7, 5.9, 5.11, 5.16 are interesting. From one point of view they are all different however each shows geographical contiguity and so we can conclude that in some sense the techniques are capturing commonalities between the time series. In all of the maps the states surrounding New York are by and large clustered with New York and also we see the mid-western state consistently co-clustered. Florida and California differ and there also seems to be a belt of states Oregon, Nevada and Arizona that are clustered together in several cases. Perhaps Nevada and Arizona make sense both being primarily desert states but only Oregon's east contains desert. However, while there is some consistency

between the clusters when it comes to producing forecasts from the clusters we see that only the SOEM based cluster provides forecast improvement over the spatial clusters in Section 5.2. This is impressive as the comparison is not fair for the SOEM based models which are restricted to three clusters, the spatial models in essence containing a unique clustering for each state.

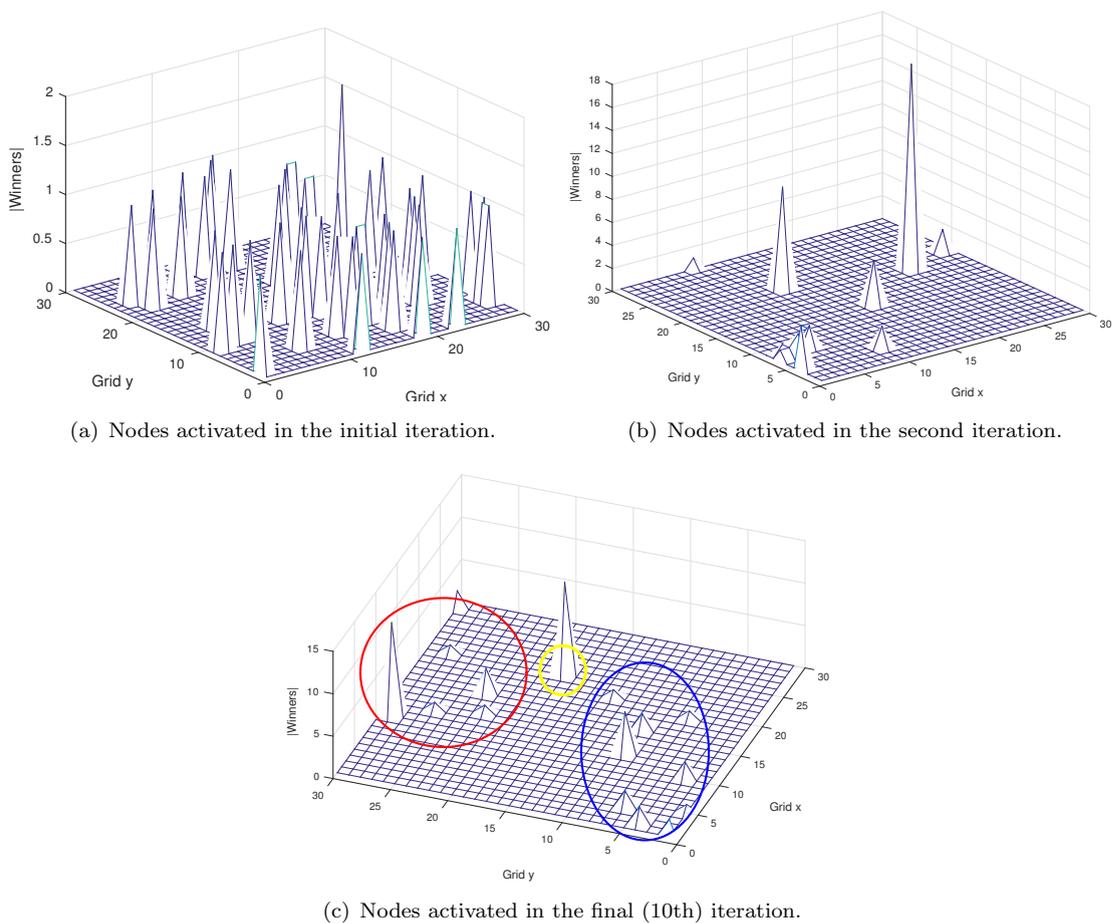


FIGURE 5.13: Distribution of winning nodes over several iterations.

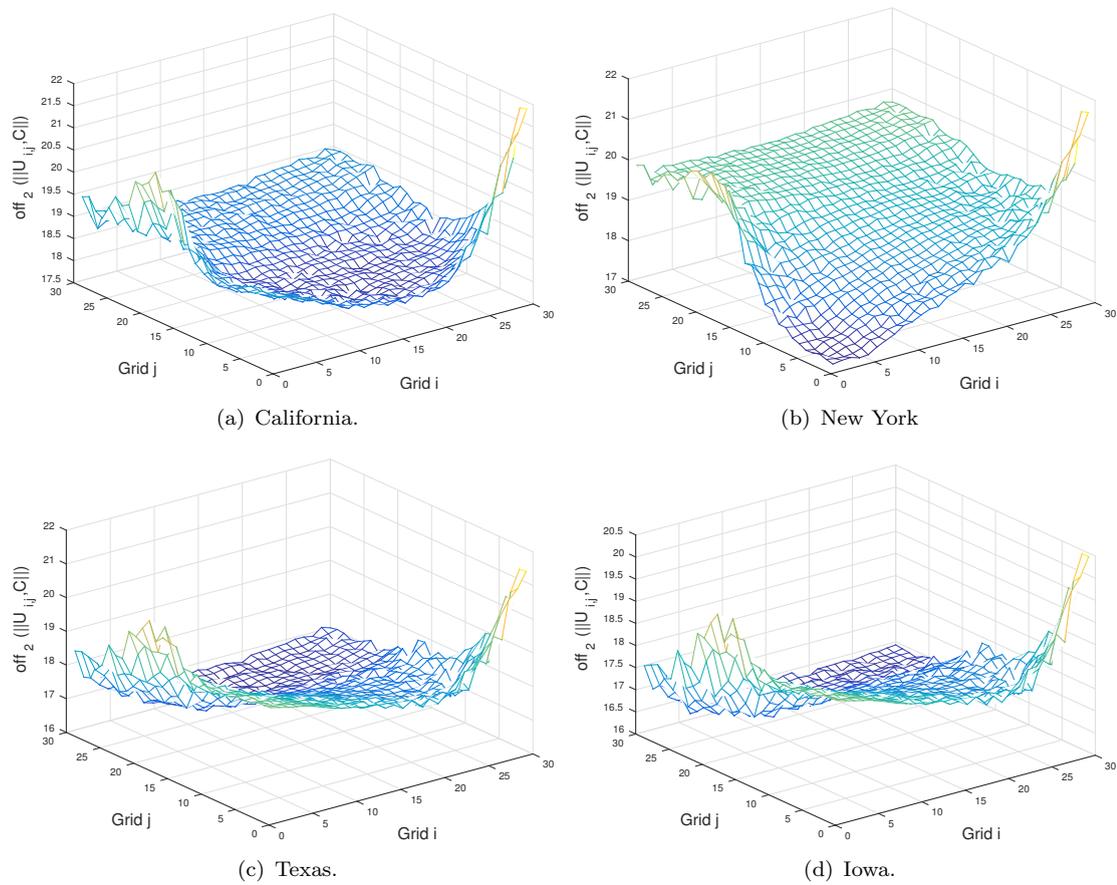


FIGURE 5.14: Distance measure, $\langle \{U_{i,j}\}, C \rangle$, for four sample American states.

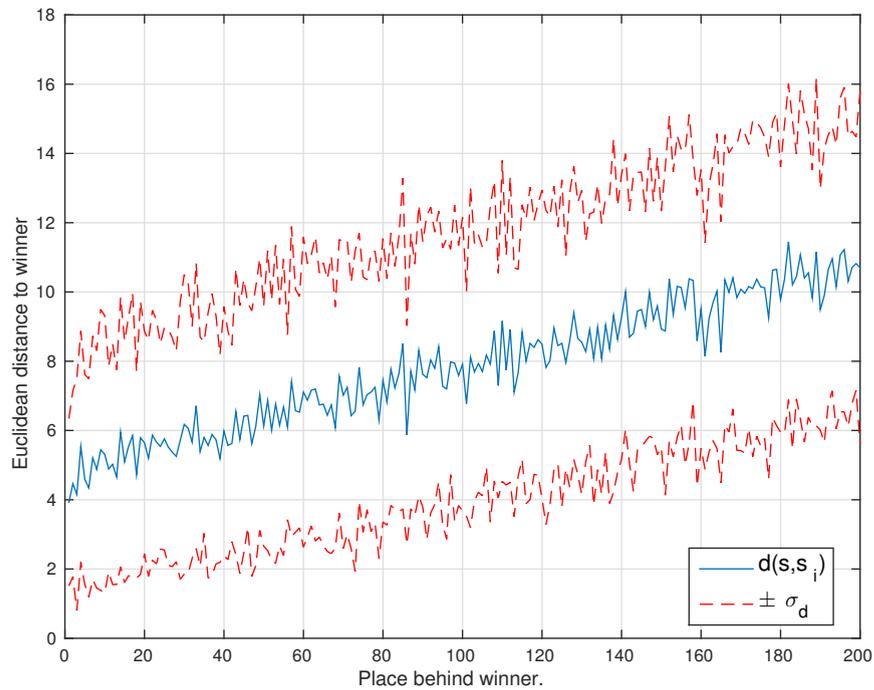


FIGURE 5.15: The x-axis is the number of places behind the winning node (i.e. 2nd place, 3rd place etc), the y-axis shows the average euclidean distance from the winning node of that node.



FIGURE 5.16: States clustered with SOEM.

Chapter 6

Conclusion

The one unifying element in this thesis is the focus on the SSA modes in multivariate time series, their relationship with each other and their evolution. SSA is based on the eigenvalue-eigenvector decomposition of the M-lag covariance matrix, therefore these modes can be expressed via the obtained eigenvectors. These eigenvectors are the core element of the SSA algorithm, in both the reconstruction and forecasting stages.

Our preliminary analysis was designed to show that we can use bootstrapping on the coefficients as well as the forecasts in SSA. The performance improvement is marginal but most interesting is perhaps Figure 3.3 in which we see that post a structural break the coefficients of the LRF are less concentrated and most worryingly this persists. SSA forecasting uses information from the eigenvectors to find the LRF coefficients. We found that the LRF is violated post break (Section 4.2.2). This outcome is also confirmed by the distribution of the coefficients which shows they are less concentrated around their mean (Section 3.3.2). In fact this uncertainty can bring more uncertainty into forecasting as well. To address this issue we proposed a state dependent coefficients in which the transition function depends on the state (Section 4.2.3).

The state dependent model was then empirically tested in both univariate and multivariate cases. For example, for the univariate case, we benchmarked against the most commonly used classical time series models such as ARIMA, ARFIMA, ETS, GARCH, SSA and bootstrap SSA (see Tables 4.4, 4.5 and 4.6), and observed most notably our proposed model, Bayesian SSA outperforms them all. The same results are also reported for the multivariate case (see Tables 4.11, 4.12 and 4.13), where the benchmarks are the VAR, VECM and MSSA models. The classical models basically suffer in the presence of a structural break but as found in (Serrano & Robles Fernandez, 2001), surprisingly, including structural break information into the model results in no significant gains. Additionally, the effects of a structural break last only as long as the break itself and the maximum effective lag of the autoregressive models, whereas for SSA it lasts longer as the trajectory matrix contains a full historical time series.

Although MBSSA outperforms MSSA, it was found that the existence of common modes between the time series (Industrial Production Indicator) perturbs the performance of separability in MSSA (Section 4.4.2.3). Therefore, we found that MBSSA-8, for example, does not always perform as good as MBSSA-2. The reason for this is that MBSSA cannot separate the modes of similar but distinct time series. Using ICA-MBSSA-8 increases the accuracy of the model significantly. ICA is just one possible preprocessing step to reduce multicollinearity. However, it would be desirable to separate time series with different modes in the first instance before modelling. However, testing all combinations of time series requires $N(2^{N-1} - 1)$ models be tested which becomes computationally complex and overly time-consuming.

In Chapter 5 we further investigated the above idea using the USA unemployment rate series. Since they are geographically distributed, information can be used to model cross-sectional spatial dependency among multiple time series. This information is incorporated into the model as an explanatory series by using two different weightings, the Inverse-Distance weight and the Distance-Boundary weights. The Moran test, Table B.2, shows that there is a statistically significant spatial interdependency in each case. In fact, the gains made by combining spatial interdependency and Bayesian MSSA and as we expected multivariate models are far better than the univariate ones (Table 5.1).

Next, it was found that the angle between the USA unemployment time series eigenvectors differs slightly, Figure 5.4. In spite of the slight differences small angles between eigenvectors can lead to significant forecast differences as shown later in Tables 5.3 and 5.4. Our first attempt was to replace the SVD used in the decomposition stage of the SSA algorithm with joint diagonalisation to include communality between multiple time series modes (Section 5.3) Surprisingly, the common eigenspace provided by JD had little effect on improving the performance of the LRF coefficients. The reason for that is those common modes between different groups of series are still supporting each other. As another alternative we viewed the covariance matrices as a multidimensional array, tensor (Section 5.4), and used a higher dimensional tensor decomposition, the Tucker transform.

The Tucker transform was found, like JD, to not be an ideal replacement for the SVD. The Tucker decomposition does not spread power equally among cores, for example, most power is always given to the first core while stronger values in the second core are given by off diagonal elements. Therefore, the Tucker transform is unable to produce a set of core slices that are diagonally dominant.

A Self organising eigenspace map was later proposed (Section 5.5) to cluster the American states into sets of non-orthogonal eigenvectors. Table 5.5 shows a gain in forecasting achieved while using the SOEM over the spatial weights.

The clustering maps of the American states is yet another interesting part of this research. Five different clustering maps resulted from analysis based on kmeans-JD,

GMM-JD, Hierarchical-JD, kmeans-Tensor and the SOEM. These can be compared interestingly with maps produced in other studies such as (Kotkin J, 2013), and (Nelson & Rae, 2016). There are differences but a general agreement in terms of the contiguity of the regions detected in our study.

This thesis also reveals that it is possible to combine clustering independently with our proposed Bayesian MSSA, and so the gains made from using MBSSA are not lost when we introduce differing reconstruction paradigms (Table 5.5).

From another angle the work in this thesis found an interesting parallel between horizontal MSSA and SSA when we have a time series with a sudden structural break (Section 4). If we consider the time series pre and post structural break to be two time series then in fact this is equivalent to horizontal MSSA.

The SDM model examined assumes a Gaussian posterior for the coefficients of the LRF model but further work involving a non-Gaussian posterior may be interesting such as via the unscented Kalman filter or the particle filter. Also this might provide us with a means to reject noise in the time series which can have an adverse effect on forecasts. It would be of interest to look at different kinds of non-linearity in the SDM (i.e. change the random walk model in Equation (4.12)). This would be interesting and in fact might align with the idea of using the unscented Kalman filter or particle filter.

There are several variants of the SOFM that could also be tried such as growing a mesh rather than apriori specifying a 30×30 grid; this is because the effect of boundary nodes having only 3 neighbours can distort the final clusters (Fritzke, 1993). It would also be interesting to derive a test for the SOEM to explore if the distribution of the eigenvector bases are sampled from the same distribution as that of the covariance matrices. For an SOFM this is easily performed using a t-test (Tatoian & Hamel, 2016) but in the SOEM case we have two sets of matrices one of which is in fact a basis so its not clear if an approach such as that in two sample tests for high dimensional covariance matrices (for example (Li & Chen, 2012)) would work. Indeed, as the SOEM is at an introductory stage it would be interesting to determine when it might converge for general matrices or under what circumstances.

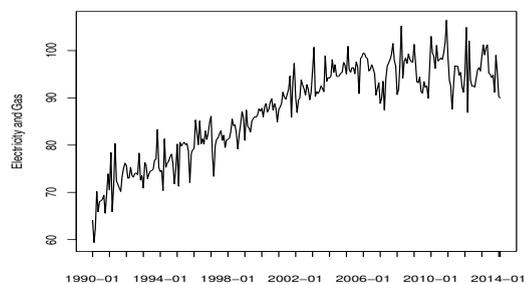
Appendix A

A.1 Granger-Newbold test

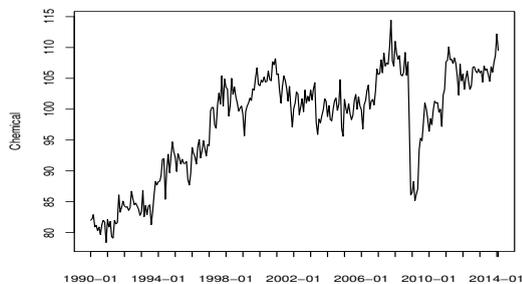
Granger and Newbold (1986) noticed an easy way to test the null of comparing forecasting errors through the following transformation. Assume $U = e_1 + e_2$, and $V = e_1 - e_2$ where e_1 and e_2 are forecast errors for model 1 and 2. Then, (U, V) has a bivariate normal distribution with parameters $E[U] = \mu_1 + \mu_2$, $E[V] = \mu_1 - \mu_2$, $Var(U) = \sigma_1^2 + \sigma_2^2 - 2\rho_e\sigma_1\sigma_2$, $Var(V) = \sigma_1^2 + \sigma_2^2 + 2\rho_e\sigma_1\sigma_2$ and $Cov(U, V) = \sigma_{UV} = \rho\sigma_U\sigma_V$. In terms of the original population, $\sigma_{UV} = \sigma_1^2 - \sigma_2^2$. If the mean squared prediction errors in the original population are equal, then the covariance in the transformed population must be zero. Then the null hypothesis in terms of the transformed population, $H_0 : Cov(U; V) = 0$. A direct way to test the null is to use the sample covariance around the population means which leads to the following statistics:

$$\frac{s_{UV}}{[\sum_{j=1}^n u_j^2 v_j^2 / n^2]^{\frac{1}{2}}} \sim N(0, 1)$$

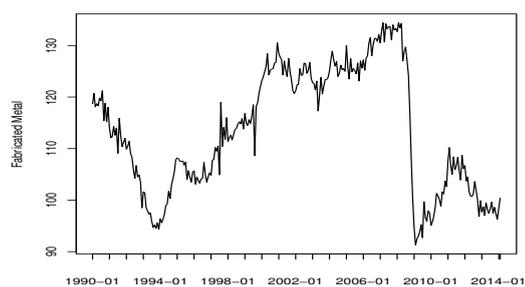
The standard precautions are called for in using a standard error to test a statistical hypothesis. It is only justified in the case where the test statistic tends to normality. For more details see (Mizrach, 1996).



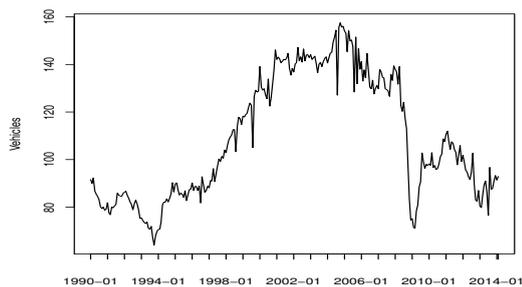
(a) Electricity and Gas



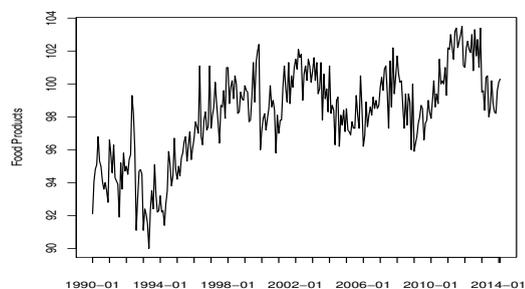
(b) Chemical



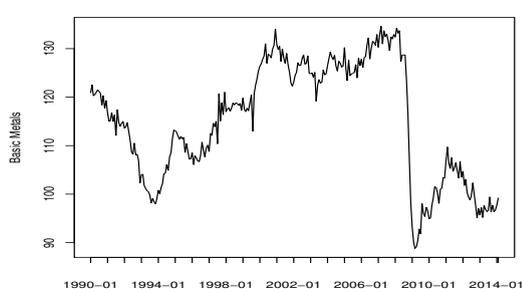
(c) Fabricate Metals



(d) Vehicle



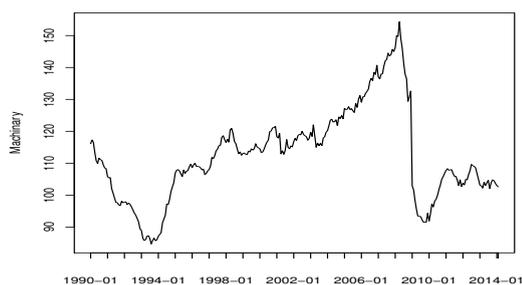
(e) Food



(f) Basic Metals



(g) Machinery



(h) Electrical Machinery

FIGURE A.1: Industrial Production Indicators for France.

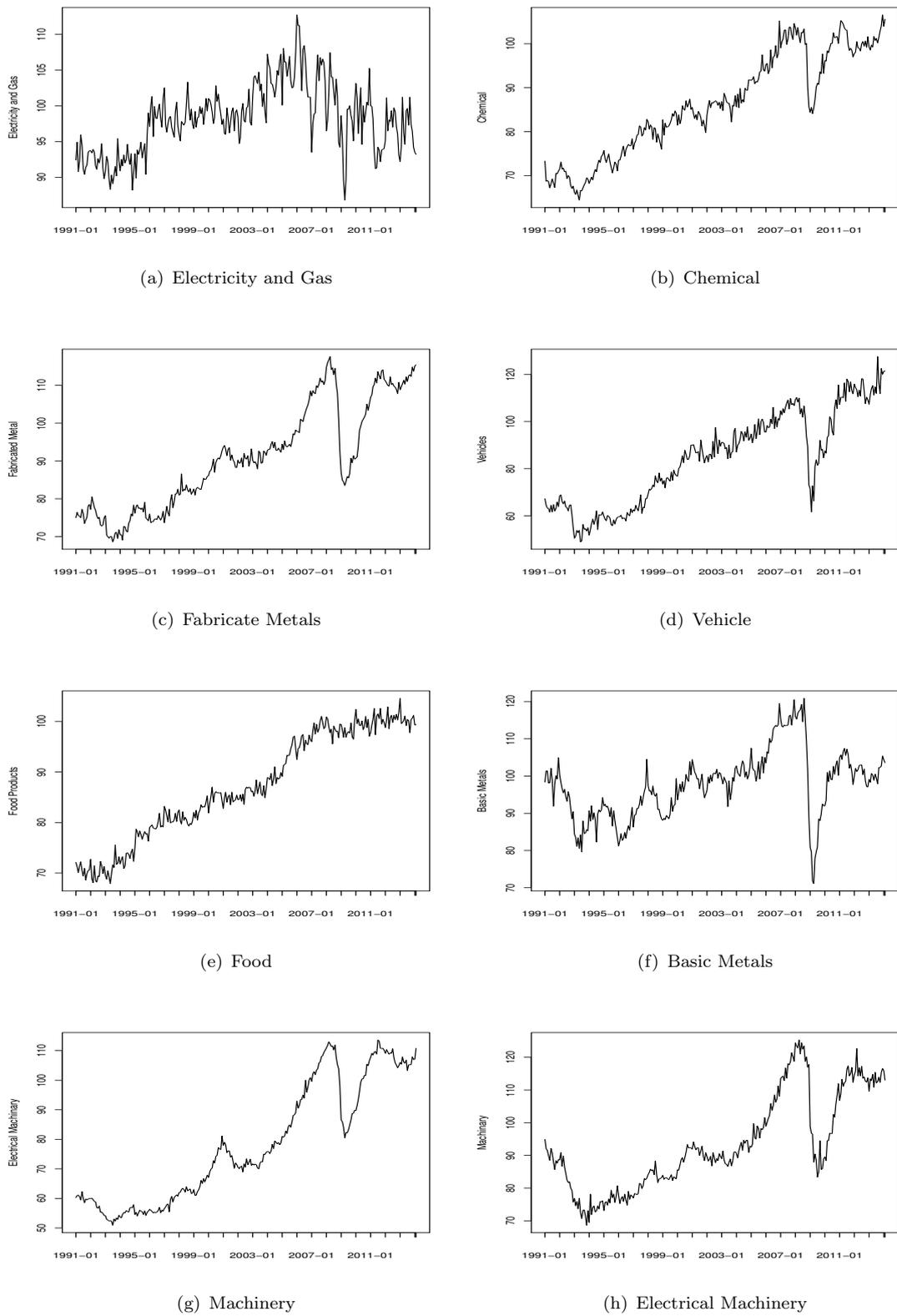
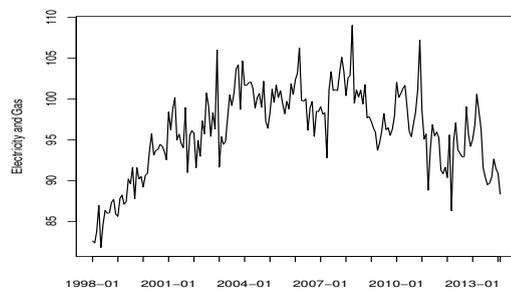
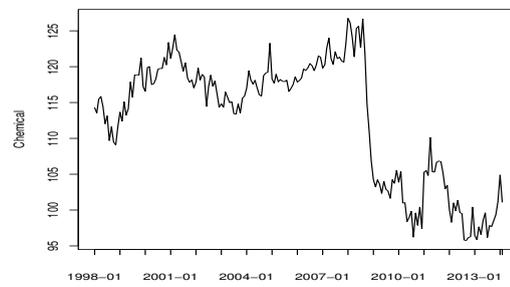


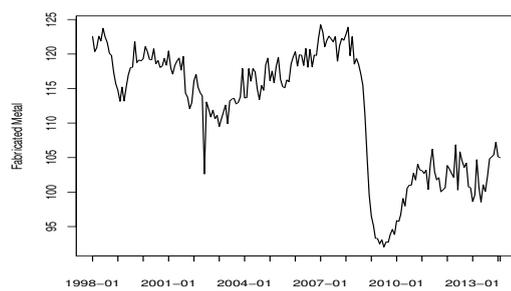
FIGURE A.2: Industrial Production Indicators for Germany.



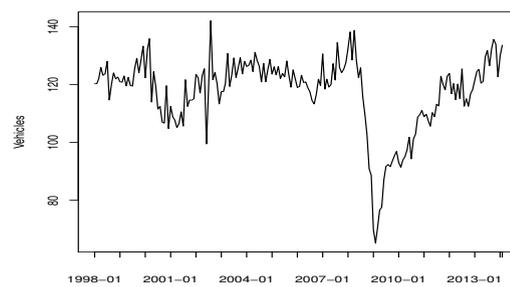
(a) Electricity and Gas



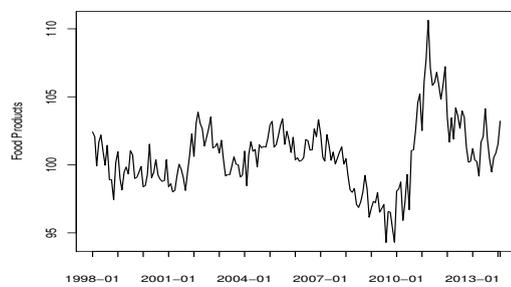
(b) Chemical



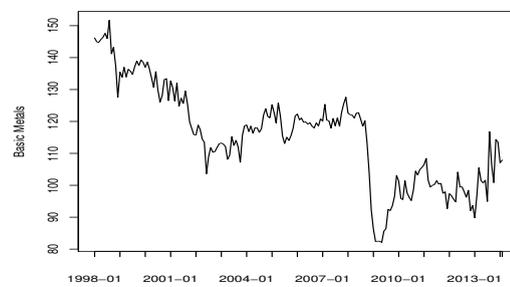
(c) Fabricate Metals



(d) Vehicle



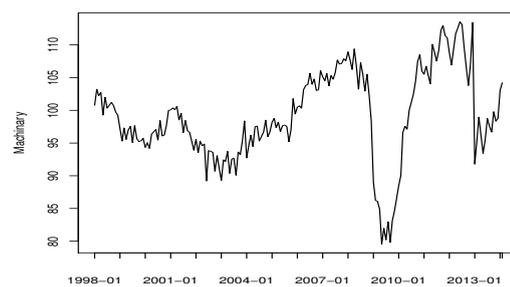
(e) Food



(f) Basic Metals



(g) Machinery



(h) Electrical Machinery

FIGURE A.3: Industrial Production Indicators for the UK.

Appendix B

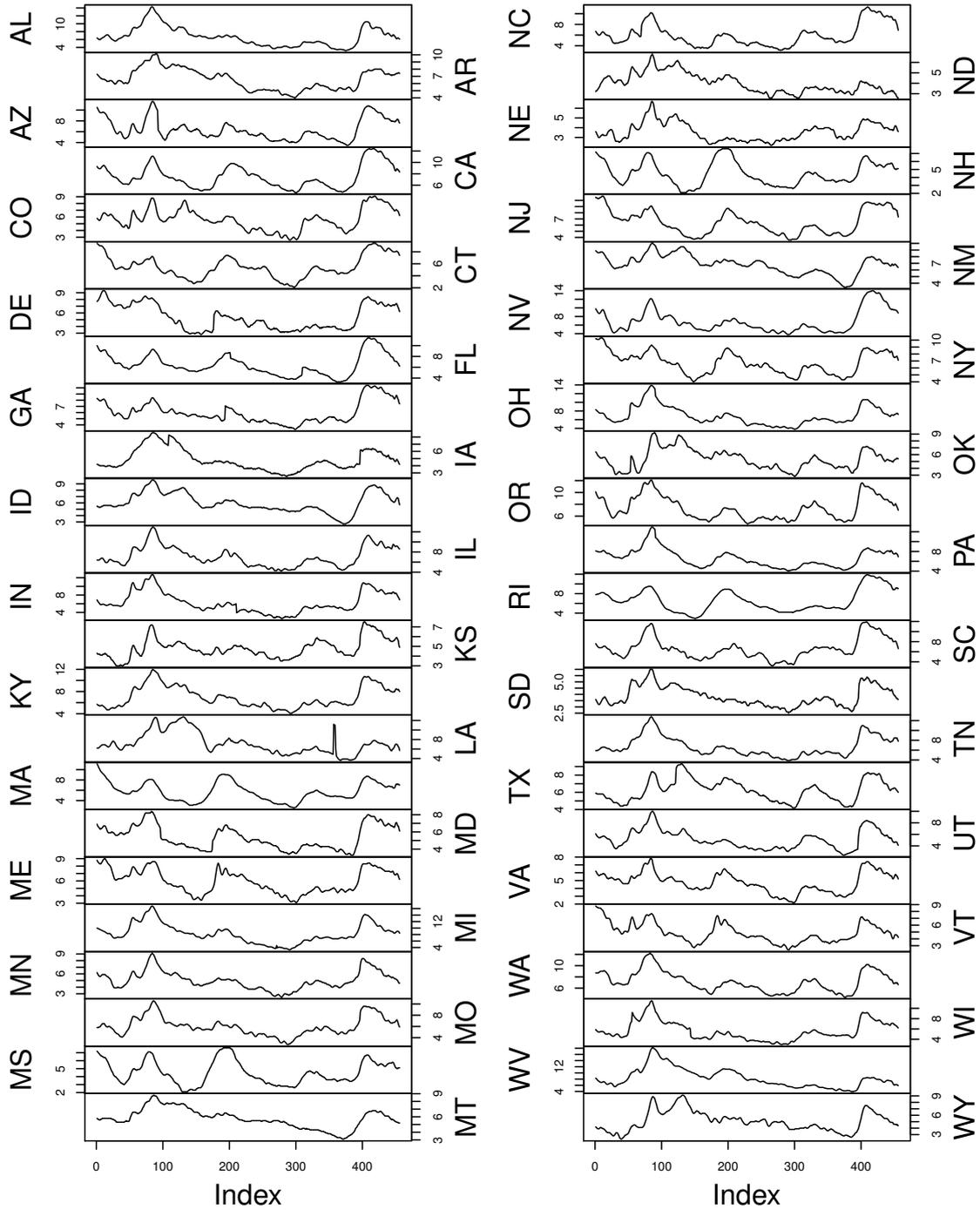


FIGURE B.1: Unemployment rates for United States.

TABLE B.1: Descriptive statistics for Unemployment Rate.

States	Mean	SD	Skewness	Kurtosis	SW	ADF	States	Mean	SD	Skewness	Kurtosis	SW	ADF
AL	6.7	2.4	0.9	0.5	0.0*	0.15	NC	6.0	2.1	1.0	0.0	0.0*	0.29
AR	6.6	1.5	0.3	-0.9	0.0*	0.61	ND	4.0	0.9	0.7	-0.3	0.0*	0.19
AZ	6.4	1.9	0.9	-0.1	0.0*	0.02	NE	3.6	0.9	0.8	0.1	0.0*	0.38
CA	7.5	2.0	0.7	-0.3	0.0*	0.02	NH	4.5	1.5	0.4	-0.8	0.0*	0.01
CO	5.6	1.6	0.3	-0.5	0.0*	0.08	NJ	6.5	1.9	0.4	-1.0	0.0*	0.04
CT	5.5	1.8	0.4	-0.5	0.0*	0.07	NM	6.8	1.4	-0.2	-0.1	0.0*	0.07
DE	5.3	1.8	0.4	-1.3	0.0*	0.50	NV	6.8	2.6	1.2	0.6	0.0*	0.08
FL	6.4	2.0	0.6	-0.1	0.0*	0.07	NY	6.7	1.6	0.2	-0.9	0.0*	0.04
GA	6.0	1.7	1.0	0.1	0.0*	0.43	OH	6.9	2.1	1.1	0.9	0.0*	0.41
IA	4.8	1.5	0.9	-0.1	0.0*	0.55	OK	5.3	1.5	0.5	-0.3	0.0*	0.28
ID	6.0	1.5	0.2	-0.4	0.0*	0.18	OR	7.3	1.9	0.7	-0.6	0.0*	0.09
IL	7.1	2.0	0.7	-0.1	0.0*	0.35	PA	6.6	1.8	0.9	1.0	0.0*	0.28
IN	6.3	2.4	0.8	-0.2	0.0*	0.58	RI	6.6	2.3	0.6	-0.5	0.0*	0.01
KS	4.8	1.0	0.7	0.3	0.0*	0.10	SC	6.6	2.0	0.9	0.3	0.0*	0.05
KY	7.0	1.9	0.7	-0.6	0.0*	0.38	SD	3.8	0.8	0.6	-0.6	0.0*	0.33
LA	7.2	2.2	0.9	0.1	0.0*	0.44	TN	6.7	2.0	1.0	0.3	0.0*	0.34
MA	5.7	1.8	0.4	-0.6	0.0*	0.02	TX	6.2	1.3	0.3	-0.6	0.0*	0.08
MD	5.4	1.4	0.4	-1.0	0.0*	0.32	UT	5.1	1.6	0.7	0.1	0.0*	0.13
ME	5.9	1.6	0.1	-1.2	0.0*	0.22	VA	4.8	1.3	0.1	-0.6	0.0*	0.23
MI	8.3	2.9	0.7	0.1	0.0*	0.37	VT	4.9	1.4	0.6	-0.3	0.0*	0.07
MN	5.0	1.3	0.7	0.4	0.0*	0.22	WA	7.1	1.8	0.7	-0.1	0.0*	0.09
MO	6.0	1.6	0.7	0.2	0.0*	0.35	WI	5.6	1.8	0.9	0.3	0.0*	0.44
MS	8.0	2.0	0.7	-0.5	0.0*	0.51	WV	8.3	3.1	1.0	0.8	0.0*	0.12
MT	5.8	1.3	0.1	-0.3	0.0*	0.49	WY	5.1	1.5	0.6	-0.1	0.0*	0.14

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.2: Local Moran test for Unemployment Rate.

States	ID			BD			States	ID			BD		
	I_i	Z_{I_i}	P-value	I_i	Z_{I_i}	P-value		I_i	Z_{I_i}	P-value	I_i	Z_{I_i}	P-value
AL	0.683	6.176	0.000	0.867	1.782	0.037	NC	0.321	2.921	0.002	-0.068	-0.089	0.535
AR	0.067	0.969	0.166	0.404	1.005	0.157	ND	0.446	5.338	0.000	3.281	5.736	0.000
AZ	-0.030	-0.088	0.535	0.272	0.585	0.279	NE	0.370	3.961	0.000	2.599	5.849	0.000
CA	-0.008	0.100	0.460	0.926	1.480	0.069	NH	0.222	1.138	0.128	1.029	1.815	0.035
CO	0.250	3.013	0.001	0.601	1.500	0.067	NJ	0.027	0.271	0.393	0.424	0.640	0.261
CT	-0.010	0.062	0.475	-0.109	-0.151	0.560	NM	0.049	0.854	0.196	0.002	0.043	0.483
DE	0.002	0.129	0.449	-0.179	-0.190	0.575	NV	-0.037	-0.126	0.550	0.573	1.188	0.117
FL	0.247	3.684	0.000	0.329	0.504	0.307	NY	0.060	0.464	0.321	0.060	0.153	0.439
GA	0.670	6.203	0.000	0.337	0.748	0.227	OH	0.310	3.194	0.001	0.873	1.942	0.026
IA	0.118	1.471	0.071	0.802	1.989	0.023	OK	0.007	0.303	0.381	0.042	0.111	0.456
ID	0.028	0.425	0.335	-0.008	0.029	0.488	OR	-0.028	-0.046	0.518	0.915	1.857	0.032
IL	0.019	0.377	0.353	-0.050	-0.063	0.525	PA	-0.014	0.051	0.480	0.272	0.650	0.258
IN	0.307	3.014	0.001	0.822	1.655	0.049	RI	0.465	1.844	0.033	0.130	0.215	0.415
KS	0.127	1.499	0.067	1.000	2.084	0.019	SC	0.344	3.055	0.001	-0.014	0.011	0.496
KY	0.456	4.709	0.000	0.588	1.474	0.070	SD	0.584	6.236	0.000	3.734	8.192	0.000
LA	0.339	3.427	0.000	1.331	2.209	0.014	TN	0.547	5.565	0.000	0.407	1.133	0.129
MA	0.437	1.708	0.044	0.202	0.508	0.306	TX	0.032	0.693	0.244	0.042	0.109	0.457
MD	0.015	0.205	0.419	0.070	0.186	0.426	UT	0.016	0.355	0.361	0.004	0.057	0.477
ME	0.320	2.053	0.020	1.586	1.634	0.051	VA	0.155	1.338	0.090	0.032	0.114	0.455
MI	0.047	0.724	0.235	0.006	0.044	0.482	VT	0.201	1.316	0.094	0.518	0.831	0.203
MN	0.265	3.134	0.001	1.713	3.582	0.000	WA	-0.042	-0.140	0.556	0.915	1.303	0.096
MO	-0.012	0.111	0.456	0.051	0.177	0.430	WI	-0.006	0.155	0.439	-0.353	-0.640	0.739
MS	0.396	3.673	0.000	1.339	2.629	0.004	WV	0.094	1.079	0.140	-0.317	-0.613	0.730
MT	0.139	1.601	0.055	0.650	1.186	0.118	WY	0.286	3.151	0.001	0.963	2.375	0.009

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.3: Bivariate post-sample forecast accuracy measures with spatial weights for UR of states AL,AR,AZ,CA,CO,CT,DE,FL.

Series	Steps	RMSE								RRMSE				
		h	SS	BS	ID-MS	BD-MS	ID-MBS	BD-MBS	Tens-MS	Tens-NBS	$\frac{ID-MS}{S}$	$\frac{ID-MBS}{BS}$	$\frac{ID-MBS}{ID-MS}$	$\frac{BD-MBS}{BD-MS}$
AL	1	0.12	0.12	0.12	0.12	0.06	0.06	0.13	0.06	1.00	0.54 *	0.52 *	0.54 *	1.02
	3	0.41	0.22	0.40	0.37	0.17	0.17	0.42	0.18	0.99	0.78 *	0.42 *	0.45 *	1.01
	6	0.82	0.46	0.81	0.78	0.35	0.35	0.96	0.44	0.98	0.76 *	0.43 *	0.45 *	1.01
	12	1.56	1.13	1.56	1.21	0.80	0.79	2.26	1.13	1.00	0.71 *	0.51 *	0.65 *	0.99
AR	1	0.06	0.07	0.06	0.07	0.05	0.05	0.07	0.05	0.98	0.76 *	0.85 *	0.78 *	0.98
	3	0.17	0.14	0.16	0.18	0.13	0.13	0.20	0.13	0.97	0.94 *	0.79 *	0.70 *	1.00
	6	0.39	0.26	0.40	0.42	0.24	0.24	0.50	0.26	1.04	0.95 *	0.61 *	0.58 *	1.00
	12	0.71	0.44	0.71	0.68	0.42	0.42	1.15	0.44	1.01	0.94 *	0.59 *	0.62 *	1.00
AZ	1	0.09	0.10	0.09	0.09	0.07	0.07	0.09	0.06	1.00	0.68 *	0.75 *	0.74 *	0.99
	3	0.27	0.18	0.26	0.26	0.17	0.17	0.27	0.15	0.98	0.97 *	0.66 *	0.66 *	0.99
	6	0.55	0.34	0.59	0.55	0.32	0.31	0.56	0.33	1.06	0.92 *	0.54 *	0.57 *	0.99
	12	1.20	0.70	1.19	1.17	0.59	0.58	1.32	0.70	0.99	0.84 *	0.50 *	0.50 *	0.99
CA	1	0.10	0.09	0.10	0.09	0.05	0.05	0.10	0.05	1.00	0.55	0.54 *	0.53 *	0.98 *
	3	0.49	0.14	0.29	0.29	0.13	0.12	0.29	0.11	0.59 *	0.89 *	0.44 *	0.43 *	0.98
	6	0.75	0.32	0.66	0.57	0.28	0.28	0.58	0.30	0.87 *	0.88 *	0.43 *	0.49 *	0.99
	12	1.13	0.84	1.16	1.11	0.68	0.68	1.30	0.83	1.02	0.81 *	0.59 *	0.61 *	0.99
CO	1	0.10	0.10	0.09	0.09	0.07	0.07	0.10	0.06	0.91	0.67 *	0.78 *	0.79 *	1.01
	3	0.31	0.18	0.30	0.30	0.18	0.19	0.33	0.16	0.98	0.99 *	0.61 *	0.61 *	1.02
	6	0.73	0.40	0.72	0.71	0.36	0.36	0.79	0.39	0.99	0.89 *	0.49 *	0.51 *	1.01
	12	1.44	0.87	1.19	1.10	0.66	0.66	1.64	0.87	0.83 *	0.76 *	0.55 *	0.60 *	1.00
CT	1	0.08	0.08	0.07	0.07	0.06	0.06	0.07	0.05	0.81	0.75 *	0.88 *	0.81 *	0.92 *
	3	0.22	0.15	0.22	0.22	0.14	0.14	0.22	0.12	1.00	0.63 *	0.63 *	0.67 *	1.06
	6	0.46	0.32	0.49	0.46	0.26	0.31	0.50	0.26	1.07	0.58 *	0.54 *	0.68 *	1.19
	12	0.94	0.63	0.85	0.93	0.62	0.63	1.04	0.62	0.91	0.66 *	0.72 *	0.67 *	1.02
DE	1	0.10	0.11	0.10	0.10	0.09	0.07	0.10	0.07	1.00	0.86 *	0.96	0.70	0.72 *
	3	0.28	0.26	0.28	0.27	0.17	0.26	0.28	0.16	1.00	0.66 *	0.61	0.93	1.50
	6	0.55	0.43	0.53	0.52	0.29	0.43	0.58	0.29	0.97	0.67 *	0.54 *	0.83 *	1.48
	12	1.33	0.69	0.97	1.26	0.57	0.68	1.25	0.57	0.73 *	0.84 *	0.59 *	0.54 *	1.19
FL	1	0.17	0.12	0.16	0.16	0.08	0.08	0.16	0.07	0.95	0.65 *	0.49 *	0.50 *	1.00
	3	0.39	0.20	0.38	0.38	0.18	0.18	0.43	0.17	0.99	0.87 *	0.46 *	0.47 *	1.00
	6	0.70	0.37	0.72	0.69	0.34	0.34	0.96	0.36	1.04	0.92 *	0.47 *	0.49 *	1.00
	12	1.36	0.81	1.30	1.10	0.69	0.69	2.86	0.81	0.95	0.85 *	0.53 *	0.63 *	1.00

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.4: Bivariate post-sample forecast accuracy measures with spatial weights
for UR of states GA,IA,ID,IL,IN,KS,KY,LA.

Series	Steps	RMSE								RRMSE				
		h	SS	BS	ID-MS	BD-MS	ID-MBS	BD-MS	Tens-MS	Tens-NBS	$\frac{ID-MS}{S}$	$\frac{ID-MBS}{BS}$	$\frac{ID-MBS}{ID-MS}$	$\frac{BD-MBS}{BD-MS}$
GA	1	0.12	0.11	0.12	0.18	0.10	0.11	0.11	0.07	0.97	0.91 *	0.86	0.60	1.07
	3	0.30	0.26	0.29	0.29	0.18	0.26	0.32	0.16	0.97	0.70 *	0.62 *	0.88 *	1.42
	6	0.56	0.45	0.56	0.56	0.33	0.45	0.60	0.32	1.01	0.73 *	0.59 *	0.80 *	1.37
	12	1.58	0.75	1.11	1.24	0.71	0.75	1.33	0.72	0.70 *	0.96 *	0.64 *	0.60 *	1.05
IA	1	0.21	0.18	0.21	0.21	0.12	0.12	0.26	0.16	1.00	0.66 *	0.56 *	0.56 *	1.00
	3	0.35	0.28	0.35	0.35	0.21	0.21	0.81	0.27	1.00	0.73 *	0.59 *	0.59 *	1.00
	6	0.49	0.39	0.51	0.49	0.29	0.29	2.67	0.38	1.04	0.75 *	0.57 *	0.60 *	1.00
	12	0.80	0.69	0.80	0.64	0.50	0.50	0.81	0.68	1.00	0.73 *	0.62 *	0.78 *	1.00
ID	1	0.08	0.09	0.08	0.08	0.06	0.06	0.08	0.06	1.00	0.63 *	0.71 *	0.71 *	1.00
	3	0.24	0.14	0.24	0.24	0.12	0.12	0.24	0.13	0.99	0.85 *	0.50 *	0.49 *	0.99
	6	0.42	0.27	0.43	0.41	0.23	0.23	0.43	0.27	1.03	0.85 *	0.54 *	0.55 *	0.99
	12	0.93	0.57	0.80	0.66	0.46	0.45	0.85	0.57	0.87 *	0.81 *	0.57 *	0.69 *	0.98
IL	1	0.12	0.11	0.11	0.11	0.07	0.07	0.12	0.06	1.00	0.64 *	0.59 *	0.59 *	1.00
	3	0.38	0.20	0.38	0.38	0.18	0.18	0.39	0.18	1.00	0.91 *	0.48 *	0.48 *	1.01
	6	0.74	0.38	0.77	0.72	0.35	0.35	0.82	0.37	1.05	0.90 *	0.45 *	0.48 *	1.00
	12	1.38	0.77	1.24	1.18	0.61	0.61	1.75	0.78	0.90	0.79 *	0.49 *	0.52 *	1.00
IN	1	0.12	0.12	0.12	0.12	0.11	0.11	0.12	0.07	1.00	0.87 *	0.88 *	0.87 *	0.99
	3	0.35	0.30	0.35	0.35	0.24	0.33	0.40	0.20	0.99	0.80 *	0.69	0.94	1.37
	6	0.74	0.50	0.72	0.72	0.47	0.56	0.95	0.49	0.98	0.94 *	0.65 *	0.79 *	1.20
	12	1.38	1.21	1.37	1.76	1.01	1.02	2.17	1.21	1.00	0.84 *	0.74 *	0.58 *	1.00
KS	1	0.12	0.10	0.12	0.12	0.07	0.06	0.12	0.07	0.98	0.69 *	0.55 *	0.53 *	0.97 *
	3	0.34	0.18	0.34	0.34	0.15	0.15	0.35	0.16	1.00	0.85 *	0.44 *	0.43 *	0.98 *
	6	0.59	0.31	0.61	0.61	0.27	0.27	0.78	0.31	1.04	0.86 *	0.44 *	0.44 *	0.99 *
	12	1.02	0.55	0.94	0.73	0.46	0.45	2.11	0.53	0.92	0.84 *	0.49 *	0.62 *	0.98 *
KY	1	0.10	0.10	0.10	0.10	0.06	0.06	0.10	0.06	1.00	0.64 *	0.65 *	0.65 *	1.00
	3	0.30	0.19	0.30	0.30	0.17	0.16	0.31	0.16	1.00	0.89 *	0.54 *	0.54 *	0.99
	6	0.64	0.41	0.68	0.63	0.35	0.35	0.74	0.40	1.06	0.87 *	0.52 *	0.55 *	0.99
	12	1.19	0.88	1.19	1.06	0.70	0.70	1.76	0.88	1.00	0.79 *	0.59 *	0.66 *	1.00
LA	1	0.79	0.51	0.79	0.79	0.32	0.31	0.79	0.42	1.00	0.62 *	0.40 *	0.39 *	0.97 *
	3	1.29	0.81	1.26	1.25	0.60	0.58	1.29	0.79	0.97	0.74 *	0.48 *	0.46 *	0.97 *
	6	1.45	1.05	1.37	1.35	0.82	0.80	1.45	1.05	0.94	0.79 *	0.60 *	0.59 *	0.96 *
	12	1.58	1.42	1.57	1.68	1.31	1.15	1.63	1.46	0.99	0.92 *	0.83 *	0.68 *	0.88 *

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.5: Bivariate post-sample forecast accuracy measures with spatial weights for UR of states MA,MD,ME,MI,MN,MO,MS,MT.

Series	Steps	RMSE								RRMSE				
		h	SS	BS	ID-MS	BD-MS	ID-MBS	BD-MS	Tens-MS	Tens-NBS	$\frac{ID-MS}{S}$	$\frac{ID-MBS}{BS}$	$\frac{ID-MBS}{ID-MS}$	$\frac{BD-MBS}{BD-MS}$
MA	1	0.08	0.08	0.08	0.08	0.05	0.05	0.08	0.05	0.99	0.61 *	0.65 *	0.64 *	0.98
	3	0.23	0.15	0.23	0.23	0.13	0.13	0.23	0.13	1.00	0.88 *	0.57 *	0.56 *	0.98
	6	0.50	0.31	0.43	0.44	0.30	0.31	0.47	0.32	0.87	0.96 *	0.70 *	0.72 *	1.04
	12	0.99	0.63	0.86	0.93	0.60	0.63	1.06	0.62	0.87	0.95	0.70 *	0.67 *	1.04
MD	1	0.08	0.09	0.09	0.09	0.07	0.08	0.09	0.06	1.10	0.78 *	0.81 *	0.83 *	1.03
	3	0.27	0.17	0.26	0.26	0.18	0.18	0.27	0.16	0.95	1.03	0.70 *	0.71 *	1.02
	6	0.52	0.28	0.47	0.48	0.29	0.32	0.58	0.29	0.89	1.02	0.62 *	0.67 *	1.10
	12	0.94	0.52	0.87	0.98	0.53	0.54	1.20	0.51	0.93	1.01	0.60 *	0.56 *	1.03
ME	1	0.08	0.09	0.08	0.08	0.07	0.07	0.09	0.06	1.00	0.78 *	0.84 *	0.87 *	1.03
	3	0.25	0.16	0.27	0.27	0.17	0.19	0.28	0.15	1.11	1.05	0.62 *	0.70 *	1.12
	6	0.48	0.32	0.51	0.51	0.33	0.38	0.61	0.33	1.06	1.02	0.65 *	0.75 *	1.16
	12	0.92	0.62	0.90	1.06	0.62	0.63	1.30	0.61	0.98	1.01	0.69 *	0.60 *	1.02
MI	1	0.10	0.12	0.10	0.09	0.08	0.08	0.11	0.06	0.99	0.64 *	0.81 *	0.83 *	1.00
	3	0.37	0.22	0.37	0.37	0.21	0.21	0.39	0.18	1.00	0.94 *	0.56 *	0.56 *	1.00
	6	0.85	0.46	0.88	0.84	0.39	0.39	0.96	0.44	1.04	0.85 *	0.45 *	0.47 *	1.00
	12	1.72	1.24	1.71	1.55	0.94	0.94	2.31	1.23	0.99	0.75 *	0.55 *	0.61 *	1.00
MN	1	0.10	0.09	0.10	0.10	0.06	0.06	0.11	0.06	1.00	0.66 *	0.55 *	0.55 *	1.00
	3	0.34	0.17	0.33	0.34	0.14	0.15	0.34	0.15	0.99	0.83 *	0.43 *	0.42 *	1.01
	6	0.66	0.38	0.65	0.66	0.30	0.30	0.81	0.37	0.99	0.79 *	0.46 *	0.46 *	1.00
	12	1.03	0.79	1.03	0.87	0.61	0.60	1.90	0.80	1.00	0.77 *	0.59 *	0.70 *	0.99
MO	1	0.17	0.12	0.16	0.16	0.07	0.07	0.16	0.07	0.98	0.60 *	0.43 *	0.43 *	1.00
	3	0.46	0.24	0.48	0.47	0.20	0.20	0.49	0.22	1.05	0.82 *	0.41 *	0.43 *	1.01
	6	0.75	0.46	0.79	0.78	0.38	0.38	1.08	0.45	1.05	0.83 *	0.48 *	0.49 *	1.00
	12	1.27	0.94	1.21	1.07	0.72	0.72	2.89	0.93	0.95	0.77 *	0.60 *	0.67 *	1.00
MS	1	0.39	0.22	0.38	0.38	0.13	0.13	0.44	0.16	0.98	0.59 *	0.34 *	0.33 *	0.98
	3	0.61	0.48	0.69	0.69	0.37	0.36	1.45	0.46	1.12	0.76 *	0.53 *	0.53 *	0.99
	6	0.98	0.75	1.05	0.97	0.58	0.58	5.07	0.75	1.07	0.78 *	0.56 *	0.60 *	0.99
	12	1.40	1.09	1.33	1.38	0.85	0.85	53.08	1.11	0.95	0.78 *	0.64 *	0.61 *	0.99
MT	1	0.06	0.05	0.06	0.06	0.05	0.05	0.06	0.04	0.96	0.86 *	0.78 *	0.76 *	0.98
	3	0.34	0.09	0.28	0.14	0.08	0.09	0.14	0.08	0.82 *	0.91 *	0.29 *	0.68 *	1.15
	6	0.75	0.18	0.39	0.24	0.18	0.18	0.28	0.18	0.52 *	0.99 *	0.46 *	0.73 *	0.97
	12	1.01	0.42	0.77	0.79	0.35	0.53	0.55	0.41	0.77 *	0.83 *	0.45 *	0.67 *	1.54

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.6: Bivariate post-sample forecast accuracy measures with spatial weights
for UR of states NC,ND,NE,NH,NJ,NM,NV,NY.

Series	Steps	RMSE								RRMSE				
		h	SS	BS	ID-MS	BD-MS	ID-MBS	BD-MS	Tens-MS	Tens-NBS	$\frac{ID-MS}{S}$	$\frac{ID-MBS}{BS}$	$\frac{ID-MBS}{ID-MS}$	$\frac{BD-MBS}{BD-MS}$
NC	1	0.10	0.11	0.10	0.10	0.07	0.07	0.11	0.06	0.99	0.63 *	0.68 *	0.67 *	0.99
	3	0.30	0.18	0.31	0.31	0.16	0.16	0.35	0.15	1.03	0.88 *	0.51 *	0.49 *	0.98
	6	0.63	0.40	0.74	0.64	0.34	0.34	0.73	0.39	1.18	0.87 *	0.46 *	0.53 *	0.99
	12	1.24	0.91	1.25	1.16	0.72	0.72	1.71	0.91	1.01	0.79 *	0.58 *	0.62 *	1.00
ND	1	0.08	0.08	0.07	0.07	0.06	0.06	0.07	0.06	0.90	0.82 *	0.86 *	0.85 *	0.98
	3	0.19	0.15	0.20	0.20	0.14	0.14	0.20	0.14	1.04	0.97 *	0.72 *	0.70 *	0.98
	6	0.32	0.24	0.41	0.32	0.24	0.24	0.45	0.23	1.26	1.01	0.59 *	0.75 *	1.01
	12	0.48	0.37	0.45	0.49	0.38	0.37	0.85	0.37	0.94	1.00	0.84 *	0.76 *	0.99
NE	1	0.07	0.08	0.07	0.07	0.06	0.06	0.08	0.06	0.99	0.76 *	0.80 *	0.79 *	0.98
	3	0.23	0.16	0.22	0.22	0.14	0.14	0.23	0.15	0.99	0.89 *	0.62 *	0.63 *	0.99
	6	0.34	0.24	0.38	0.34	0.22	0.22	0.45	0.23	1.11	0.92 *	0.58 *	0.64 *	1.00
	12	0.66	0.42	0.51	0.47	0.36	0.36	0.83	0.42	0.78 *	0.84 *	0.70 *	0.75 *	0.99
NH	1	0.07	0.08	0.07	0.07	0.05	0.05	0.07	0.05	1.00	0.68 *	0.74 *	0.73 *	0.98
	3	0.21	0.15	0.21	0.21	0.14	0.14	0.21	0.15	1.00	0.91 *	0.68 *	0.67 *	0.99
	6	0.44	0.31	0.46	0.44	0.30	0.30	0.48	0.31	1.04	0.97	0.66 *	0.69 *	1.00
	12	0.92	0.52	0.83	0.84	0.53	0.60	1.08	0.51	0.90	1.03	0.64 *	0.72 *	1.13
NJ	1	0.09	0.09	0.09	0.09	0.06	0.06	0.09	0.06	1.00	0.63 *	0.65 *	0.65 *	1.00
	3	0.24	0.15	0.24	0.24	0.15	0.15	0.25	0.13	1.00	0.94 *	0.60 *	0.60 *	1.01
	6	0.54	0.35	0.56	0.55	0.32	0.33	0.57	0.35	1.04	0.92 *	0.58 *	0.59 *	1.01
	12	1.01	0.81	1.08	1.04	0.70	0.70	1.39	0.82	1.08	0.86 *	0.64 *	0.68 *	1.00
NM	1	0.08	0.07	0.08	0.08	0.05	0.05	0.08	0.05	1.00	0.75 *	0.70 *	0.70 *	1.00
	3	0.19	0.12	0.19	0.19	0.11	0.13	0.19	0.12	0.97	0.90 *	0.58 *	0.69 *	1.17
	6	0.37	0.26	0.37	0.37	0.24	0.26	0.38	0.26	1.02	0.93 *	0.64 *	0.71 *	1.09
	12	0.75	0.55	0.70	0.77	0.50	0.50	0.78	0.55	0.94	0.91 *	0.71 *	0.66 *	1.01
NV	1	0.10	0.10	0.08	0.08	0.06	0.06	0.08	0.05	0.80	0.59 *	0.79 *	0.77 *	0.98
	3	0.25	0.17	0.25	0.25	0.16	0.16	0.26	0.13	1.00	0.95 *	0.62 *	0.61 *	0.99
	6	0.63	0.36	0.69	0.63	0.33	0.33	0.63	0.34	1.09	0.93 *	0.48 *	0.53 *	0.99
	12	1.39	0.81	1.39	1.32	0.66	0.66	1.42	0.80	1.00	0.81 *	0.47 *	0.50 *	1.00
NY	1	0.09	0.09	0.09	0.09	0.06	0.06	0.09	0.06	1.00	0.70 *	0.70 *	0.69 *	0.98
	3	0.28	0.16	0.28	0.28	0.15	0.16	0.28	0.14	1.00	0.97 *	0.54 *	0.57 *	1.04
	6	0.58	0.33	0.62	0.57	0.32	0.32	0.60	0.33	1.06	0.97 *	0.52 *	0.55 *	0.99
	12	1.07	0.68	0.98	1.06	0.61	0.61	1.34	0.68	0.92 *	0.90 *	0.62 *	0.57 *	1.00

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.7: Bivariate post-sample forecast accuracy measures with spatial weights for UR of states OH,OK,OR,PA,RI,SC,SD,TN.

Series	Steps	RMSE								RRMSE				
		h	SS	BS	ID-MS	BD-MS	ID-MBS	BD-MS	Tens-MS	Tens-NBS	$\frac{ID-MS}{S}$	$\frac{ID-MBS}{BS}$	$\frac{ID-MBS}{ID-MS}$	$\frac{BD-MBS}{BD-MS}$
OH	1	0.09	0.10	0.09	0.09	0.06	0.06	0.09	0.06	0.96	0.65 *	0.69 *	0.69 *	1.00
	3	0.27	0.16	0.28	0.28	0.16	0.16	0.30	0.14	1.01	0.97 *	0.57 *	0.56 *	1.01
	6	0.60	0.35	0.60	0.59	0.32	0.32	0.70	0.35	1.01	0.91 *	0.53 *	0.55 *	1.00
	12	1.14	0.80	1.08	0.96	0.64	0.64	1.58	0.81	0.95	0.80 *	0.59 *	0.67 *	1.00
OK	1	0.10	0.10	0.09	0.09	0.08	0.07	0.10	0.07	0.98	0.77 *	0.82 *	0.78 *	0.95 *
	3	0.26	0.19	0.26	0.25	0.19	0.18	0.31	0.17	0.98	1.02	0.75 *	0.72 *	0.95 *
	6	0.57	0.36	0.55	0.55	0.35	0.34	0.72	0.35	0.96	0.99	0.65 *	0.61 *	0.96 *
	12	1.02	0.68	0.95	0.87	0.60	0.57	1.57	0.67	0.93	0.88 *	0.63 *	0.65 *	0.95 *
OR	1	0.10	0.12	0.10	0.10	0.06	0.06	0.11	0.06	1.01	0.51 *	0.58 *	0.58 *	1.00
	3	0.38	0.21	0.38	0.38	0.17	0.17	0.40	0.18	1.00	0.79 *	0.44 *	0.44 *	1.00
	6	0.78	0.49	0.94	0.77	0.40	0.40	0.97	0.48	1.22	0.80 *	0.42 *	0.52 *	1.00
	12	1.51	1.25	1.50	1.21	0.87	0.87	2.34	1.23	0.99	0.70 *	0.58 *	0.72 *	1.00
PA	1	0.09	0.08	0.09	0.09	0.06	0.06	0.09	0.06	1.00	0.68 *	0.62 *	0.62 *	1.00
	3	0.25	0.15	0.25	0.25	0.13	0.13	0.25	0.14	1.02	0.85 *	0.51 *	0.51 *	1.00
	6	0.42	0.29	0.46	0.42	0.26	0.26	0.48	0.29	1.07	0.89 *	0.57 *	0.62 *	1.00
	12	0.78	0.55	0.78	0.72	0.49	0.49	1.01	0.56	0.99	0.89 *	0.63 *	0.68 *	1.00
RI	1	0.08	0.08	0.08	0.08	0.05	0.05	0.08	0.05	1.00	0.57 *	0.57 *	0.57 *	1.00
	3	0.23	0.14	0.23	0.23	0.11	0.11	0.23	0.11	0.99	0.82 *	0.49 *	0.49 *	1.01
	6	0.47	0.25	0.46	0.46	0.25	0.26	0.49	0.25	0.99	0.99 *	0.55 *	0.55 *	1.01
	12	0.86	0.57	0.87	0.84	0.56	0.62	0.96	0.58	1.01	0.98	0.64 *	0.73 *	1.10
SC	1	0.10	0.11	0.10	0.10	0.07	0.07	0.11	0.06	1.00	0.61 *	0.67 *	0.67 *	0.99
	3	0.37	0.19	0.37	0.36	0.18	0.18	0.38	0.16	1.00	0.97	0.50 *	0.50 *	0.98
	6	0.74	0.38	0.77	0.71	0.37	0.37	0.83	0.37	1.03	0.97	0.48 *	0.51 *	0.99
	12	1.43	1.00	1.44	1.21	0.78	0.79	2.03	0.97	1.01	0.78 *	0.55 *	0.65 *	1.01
SD	1	0.13	0.09	0.12	0.12	0.06	0.06	0.12	0.07	0.92	0.68 *	0.49 *	0.48 *	1.00
	3	0.31	0.17	0.31	0.31	0.13	0.13	0.35	0.16	1.00	0.77 *	0.44 *	0.44 *	1.01
	6	0.48	0.29	0.51	0.48	0.23	0.23	0.83	0.28	1.07	0.77 *	0.44 *	0.47 *	1.00
	12	0.75	0.53	0.75	0.61	0.40	0.40	2.54	0.52	1.00	0.74 *	0.53 *	0.65 *	1.00
TN	1	0.11	0.11	0.10	0.10	0.06	0.06	0.11	0.07	0.99	0.58 *	0.61 *	0.61 *	1.00
	3	0.31	0.21	0.31	0.31	0.17	0.17	0.36	0.18	1.00	0.79 *	0.54 *	0.54 *	1.00
	6	0.74	0.46	0.75	0.72	0.36	0.36	0.92	0.45	1.01	0.77 *	0.48 *	0.49 *	1.00
	12	1.45	0.96	1.40	1.14	0.74	0.74	2.12	0.97	0.96	0.77 *	0.53 *	0.65 *	1.00

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.8: Bivariate post-sample forecast accuracy measures with spatial weights
for UR of states TX,UT,VA,VT,WA,WI,WV,WY.

Series	Steps	RMSE								RRMSE				
		h	SS	BS	ID-MS	BD-MS	ID-MBS	BD-MS	Tens-MS	Tens-NBS	$\frac{ID-MS}{S}$	$\frac{ID-MBS}{BS}$	$\frac{ID-MBS}{ID-MS}$	$\frac{BD-MBS}{BD-MS}$
TX	1	0.07	0.08	0.07	0.07	0.06	0.05	0.07	0.05	1.00	0.74 *	0.77 *	0.74 *	0.96 *
	3	0.19	0.13	0.20	0.20	0.13	0.13	0.21	0.12	1.05	1.02	0.65 *	0.64 *	0.95
	6	0.39	0.27	0.48	0.42	0.26	0.26	0.47	0.27	1.24	0.97	0.54 *	0.61 *	0.99
	12	0.69	0.57	0.78	0.78	0.51	0.48	1.06	0.57	1.12	0.89 *	0.65 *	0.61 *	0.94 *
UT	1	0.24	0.18	0.23	0.22	0.11	0.11	0.22	0.14	0.95	0.60 *	0.48 *	0.49 *	1.00
	3	0.61	0.29	0.61	0.61	0.20	0.20	0.72	0.24	1.00	0.68 *	0.33 *	0.33 *	1.01
	6	0.99	0.56	0.99	0.99	0.41	0.42	1.85	0.53	0.99	0.74 *	0.42 *	0.42 *	1.00
	12	1.59	1.03	1.65	1.26	0.72	0.72	9.10	1.01	1.04	0.70 *	0.44 *	0.57 *	1.00
VA	1	0.09	0.09	0.09	0.09	0.06	0.06	0.09	0.06	1.01	0.68 *	0.71 *	0.72 *	1.02
	3	0.27	0.17	0.27	0.27	0.16	0.16	0.28	0.15	1.00	0.94 *	0.57 *	0.59 *	1.03
	6	0.51	0.33	0.50	0.50	0.31	0.31	0.61	0.34	0.98	0.91 *	0.61 *	0.62 *	1.02
	12	0.89	0.66	0.88	0.89	0.58	0.58	1.33	0.65	0.99	0.88 *	0.66 *	0.66 *	1.01
VT	1	0.09	0.09	0.09	0.09	0.07	0.07	0.10	0.07	1.00	0.77 *	0.77 *	0.78 *	1.00
	3	0.32	0.18	0.32	0.32	0.17	0.18	0.34	0.17	1.00	0.96	0.52 *	0.56 *	1.07
	6	0.58	0.34	0.61	0.59	0.34	0.36	0.75	0.34	1.06	1.00	0.56 *	0.61 *	1.06
	12	0.86	0.71	0.91	0.92	0.65	0.66	1.40	0.71	1.06	0.92 *	0.72 *	0.71 *	1.01
WA	1	0.09	0.10	0.09	0.09	0.06	0.06	0.10	0.06	1.00	0.59 *	0.64 *	0.64 *	1.00
	3	0.31	0.17	0.31	0.31	0.14	0.14	0.32	0.15	1.00	0.82 *	0.45 *	0.45 *	1.00
	6	0.67	0.38	0.81	0.64	0.33	0.33	0.75	0.38	1.21	0.85 *	0.40 *	0.51 *	1.00
	12	1.15	0.84	1.18	0.94	0.65	0.65	1.70	0.83	1.03	0.77 *	0.55 *	0.69 *	1.00
WI	1	0.09	0.10	0.08	0.08	0.06	0.06	0.09	0.06	0.93	0.63 *	0.77 *	0.76 *	0.98 *
	3	0.27	0.19	0.25	0.25	0.17	0.17	0.31	0.16	0.94	0.87 *	0.66 *	0.66 *	1.00
	6	0.62	0.40	0.60	0.60	0.33	0.33	0.79	0.39	0.97	0.85 *	0.55 *	0.56 *	1.00
	12	1.18	0.93	1.17	1.01	0.68	0.68	1.89	0.91	0.99	0.73 *	0.58 *	0.68 *	1.00
WV	1	0.10	0.10	0.10	0.10	0.07	0.07	0.10	0.07	1.00	0.71 *	0.70 *	0.70 *	0.97 *
	3	0.34	0.18	0.34	0.33	0.17	0.17	0.35	0.16	1.00	0.96	0.51 *	0.50 *	0.97 *
	6	0.71	0.33	0.71	0.67	0.33	0.32	0.80	0.32	1.00	1.01	0.46 *	0.48 *	0.97 *
	12	1.20	0.64	1.20	1.17	0.60	0.58	1.63	0.62	1.00	0.93 *	0.50 *	0.50 *	0.97 *
WY	1	0.07	0.08	0.07	0.07	0.07	0.07	0.18	0.05	1.00	0.85 *	0.97	0.93	0.96 *
	3	0.22	0.16	0.22	0.22	0.15	0.18	0.23	0.14	0.99	0.96	0.68 *	0.81 *	1.18
	6	0.51	0.32	0.50	0.50	0.31	0.33	0.55	0.32	0.97	0.96	0.63 *	0.67 *	1.07
	12	1.01	0.62	0.97	0.94	0.57	0.55	1.23	0.61	0.96	0.91 *	0.59 *	0.58 *	0.96 *

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.9: Multivariate post-sample forecast accuracy for clustered states with JD for UR of states AL,AR,AZ,CA,CO,CT,DE,FL.

Series	Steps	RMSE						RRMSE			
		MBS	km-MBS	km-MS	GMM-MBS	GMM-MS	Hier-MBS	Hier-MS	$\frac{km-MBS}{MBS}$	$\frac{GMM-MBS}{MBS}$	$\frac{Hier-MBS}{MBS}$
AL	1	0.07	0.08	0.13	0.07	0.13	0.07	0.13	1.10	1.00	1.00
	3	0.20	0.21	0.41	0.20	0.41	0.20	0.41	1.07	1.00	1.00
	6	0.45	0.47	0.94	0.45	0.94	0.45	0.94	1.04	1.00	1.00
	12	1.08	1.00	2.15	1.08	2.16	1.08	2.17	0.92	* 1.00	1.00
AR	1	0.06	0.06	0.07	0.06	0.07	0.06	0.07	1.13	1.00	1.00
	3	0.14	0.16	0.20	0.14	0.20	0.14	0.20	1.18	1.00	1.00
	6	0.27	0.31	0.50	0.27	0.51	0.27	0.50	1.14	1.00	1.00
	12	0.44	0.50	1.16	0.43	1.18	0.44	1.16	1.15	1.00	1.00
AZ	1	0.07	0.07	0.09	0.07	0.09	0.09	0.09	0.97	* 0.97	* 1.30
	3	0.18	0.18	0.27	0.18	0.27	0.20	0.22	0.97	* 0.97	* 1.09
	6	0.37	0.36	0.56	0.36	0.57	0.41	0.43	0.97	* 0.97	* 1.10
	12	0.74	0.72	1.33	0.71	1.34	0.79	0.81	0.97	* 0.97	* 1.07
CA	1	0.06	0.07	0.10	0.06	0.10	0.06	0.10	1.05	1.00	1.00
	3	0.16	0.16	0.29	0.16	0.29	0.16	0.29	1.05	1.00	1.00
	6	0.36	0.37	0.58	0.36	0.58	0.36	0.58	1.05	1.00	1.00
	12	0.82	0.83	1.29	0.81	1.30	0.82	1.29	1.01	1.00	1.00
CO	1	0.07	0.07	0.10	0.06	0.10	0.07	0.10	0.98	0.86	* 1.00
	3	0.21	0.21	0.33	0.14	0.33	0.21	0.33	0.98	0.68	* 1.00
	6	0.45	0.44	0.78	0.29	0.78	0.45	0.78	0.98	0.65	* 1.00
	12	0.78	0.79	1.62	0.61	1.62	0.78	1.61	1.00	0.78	* 1.00
CT	1	0.06	0.06	0.07	0.06	0.07	0.06	0.07	0.87	* 0.87	* 1.01
	3	0.18	0.15	0.22	0.15	0.22	0.19	0.22	0.83	* 0.83	* 1.08
	6	0.38	0.31	0.49	0.31	0.49	0.38	0.49	0.81	* 0.81	* 0.99
	12	0.75	0.68	1.02	0.68	1.02	0.77	1.02	0.91	* 0.91	* 1.04
DE	1	0.08	0.08	0.10	0.08	0.10	0.08	0.10	1.02	1.00	1.00
	3	0.17	0.18	0.28	0.17	0.28	0.17	0.28	1.04	1.00	1.00
	6	0.34	0.36	0.59	0.34	0.59	0.34	0.59	1.04	1.00	1.00
	12	0.57	0.57	1.28	0.57	1.28	0.57	1.28	1.01	1.00	1.00
FL	1	0.09	0.09	0.15	0.08	0.15	0.09	0.15	1.00	0.88	* 1.00
	3	0.21	0.21	0.39	0.20	0.39	0.21	0.39	1.00	0.94	* 1.00
	6	0.43	0.43	0.77	0.40	0.76	0.43	0.76	1.00	0.94	* 1.00
	12	0.81	0.81	1.81	0.79	1.61	0.81	1.63	1.00	0.97	* 1.00

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.10: Multivariate post-sample forecast accuracy for clustered states with
 JD for UR of states GA,IA,ID,IL,IN,KS,KY,LA.

Series	Steps	RMSE							RRMSE		
		MBS	km-MBS	km-MS	GMM-MBS	GMM-MS	Hier-MBS	Hier-MS	$\frac{km-MBS}{MBS}$	$\frac{GMM-MBS}{MBS}$	$\frac{Hier-MBS}{MBS}$
GA	1	0.08	0.08	0.10	0.07	0.10	0.08	0.10	1.00	0.93	* 1.00
	3	0.19	0.19	0.32	0.18	0.32	0.19	0.32	1.00	0.93	* 1.00
	6	0.37	0.37	0.60	0.34	0.59	0.37	0.60	1.00	0.93	* 1.00
	12	0.74	0.73	1.38	0.74	1.28	0.74	1.33	1.00	1.00	* 1.00
IA	1	0.16	0.16	0.23	0.16	0.23	0.16	0.23	1.00	1.00	1.00
	3	0.28	0.28	0.55	0.28	0.55	0.28	0.55	1.00	1.00	1.00
	6	0.39	0.39	0.94	0.39	0.93	0.39	0.93	1.00	1.00	1.00
	12	0.68	0.68	1.93	0.68	1.91	0.68	1.90	1.00	1.00	1.00
ID	1	0.06	0.06	0.08	0.06	0.08	0.06	0.08	1.01	1.00	1.00
	3	0.14	0.14	0.24	0.14	0.24	0.14	0.24	1.02	1.00	1.00
	6	0.29	0.29	0.42	0.29	0.43	0.29	0.43	1.02	1.00	1.00
	12	0.60	0.61	0.83	0.60	0.93	0.60	0.85	1.03	1.00	1.00
IL	1	0.07	0.07	0.12	0.07	0.12	0.07	0.12	1.01	1.00	1.00
	3	0.20	0.20	0.38	0.19	0.38	0.20	0.38	1.01	1.00	1.00
	6	0.39	0.39	0.81	0.39	0.81	0.39	0.81	1.01	1.00	1.00
	12	0.78	0.78	1.72	0.78	1.75	0.78	1.72	1.00	1.00	1.00
IN	1	0.08	0.08	0.12	0.07	0.12	0.08	0.12	1.02	0.91	* 1.00
	3	0.22	0.22	0.40	0.19	0.40	0.22	0.40	1.01	0.89	* 1.00
	6	0.49	0.50	0.95	0.52	0.95	0.49	0.95	1.00	1.05	1.00
	12	1.17	1.15	2.17	1.07	2.18	1.17	2.15	0.98	0.91	* 1.00
KS	1	0.08	0.08	0.12	0.08	0.12	0.08	0.12	1.02	1.00	1.00
	3	0.19	0.20	0.33	0.19	0.33	0.19	0.33	1.02	1.00	1.00
	6	0.34	0.35	0.70	0.34	0.70	0.34	0.70	1.02	1.00	1.00
	12	0.63	0.65	1.50	0.63	1.50	0.63	1.49	1.03	1.00	1.00
KY	1	0.07	0.07	0.10	0.07	0.10	0.07	0.10	1.00	1.08	1.00
	3	0.17	0.17	0.31	0.19	0.31	0.17	0.31	1.00	1.09	1.00
	6	0.41	0.41	0.73	0.44	0.73	0.41	0.74	1.00	1.07	1.00
	12	0.86	0.86	1.69	0.82	1.65	0.86	1.71	1.00	0.94	1.00
LA	1	0.44	0.44	0.89	0.44	0.85	0.44	0.83	1.00	1.00	1.00
	3	0.81	0.81	3.00	0.81	2.57	0.81	2.40	1.00	1.00	1.00
	6	1.08	1.08	8.10	1.08	5.24	1.08	4.37	1.00	1.00	1.00
	12	1.66	1.66	54.39	1.66	16.28	1.66	9.78	1.00	1.00	1.00

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.11: Multivariate post-sample forecast accuracy for clustered states with
 JD for UR of states MA,MD,ME,MI,MN,MO,MS,MT.

Series	Steps	RMSE						RRMSE			
		MBS	km-MBS	km-MS	GMM-MBS	GMM-MS	Hier-MBS	Hier-MS	$\frac{km-MBS}{MBS}$	$\frac{GMM-MBS}{MBS}$	$\frac{Hier-MBS}{MBS}$
MA	1	0.07	0.06	0.08	0.06	0.08	0.06	0.08	0.90	* 0.90	* 0.86 *
	3	0.21	0.18	0.23	0.18	0.23	0.16	0.23	0.84	* 0.84	* 0.78 *
	6	0.42	0.38	0.47	0.38	0.47	0.36	0.47	0.91	* 0.91	* 0.86 *
	12	0.70	0.63	1.04	0.63	1.05	0.61	1.04	0.90	* 0.90	* 0.87 *
MD	1	0.07	0.07	0.09	0.07	0.09	0.07	0.09	1.02	1.00	1.00
	3	0.18	0.18	0.27	0.17	0.27	0.18	0.27	1.03	0.98	1.00
	6	0.33	0.34	0.58	0.31	0.58	0.33	0.58	1.04	0.95	1.00
	12	0.55	0.57	1.22	0.53	1.22	0.55	1.21	1.04	0.96	1.00
ME	1	0.07	0.07	0.09	0.07	0.09	0.07	0.09	1.02	1.03	1.00
	3	0.18	0.19	0.28	0.13	0.28	0.18	0.28	1.04	0.74	* 1.00
	6	0.35	0.35	0.62	0.32	0.62	0.35	0.62	1.01	0.91	* 1.00
	12	0.57	0.55	1.34	0.54	1.36	0.56	1.34	0.98	0.95	* 1.00
MI	1	0.07	0.07	0.11	0.07	0.11	0.07	0.11	1.01	1.00	1.00
	3	0.20	0.20	0.39	0.19	0.39	0.20	0.39	1.01	0.96	* 1.00
	6	0.46	0.46	0.95	0.44	0.95	0.46	0.95	1.00	0.96	* 1.00
	12	1.25	1.24	2.30	1.22	2.30	1.25	2.28	0.99	0.97	* 1.00
MN	1	0.06	0.06	0.11	0.06	0.11	0.06	0.11	1.02	0.96	* 1.00
	3	0.17	0.17	0.33	0.18	0.33	0.17	0.33	1.01	1.08	1.00
	6	0.38	0.38	0.76	0.37	0.76	0.38	0.76	1.00	0.97	* 1.00
	12	0.78	0.77	1.64	0.77	1.63	0.78	1.62	0.98	0.98	* 1.00
MO	1	0.08	0.08	0.16	0.08	0.16	0.08	0.16	1.00	1.00	1.00
	3	0.22	0.22	0.46	0.22	0.46	0.22	0.46	1.00	1.00	1.00
	6	0.46	0.46	0.93	0.46	0.94	0.46	0.94	1.00	1.00	1.00
	12	0.90	0.90	2.06	0.90	1.98	0.90	1.97	1.00	1.00	1.00
MS	1	0.17	0.19	0.38	0.19	0.38	0.24	0.38	1.14	1.14	1.40
	3	0.48	0.51	0.97	0.51	0.96	0.55	0.99	1.05	1.05	1.14
	6	0.76	0.80	1.90	0.80	1.87	0.87	1.97	1.05	1.05	1.14
	12	1.09	1.13	3.93	1.13	3.74	1.19	4.32	1.04	1.04	1.10
MT	1	0.04	0.04	0.06	0.04	0.06	0.04	0.06	1.00	1.00	1.00
	3	0.09	0.09	0.14	0.09	0.14	0.09	0.14	1.01	1.00	1.00
	6	0.20	0.20	0.28	0.20	0.28	0.20	0.28	1.01	1.00	1.00
	12	0.44	0.44	0.55	0.44	0.56	0.44	0.55	1.01	1.00	1.00

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.12: Multivariate post-sample forecast accuracy for clustered states with
JD for UR of states NC,ND,NE,NH,NJ,NM,NV,NY.

Series	Steps	RMSE						RRMSE			
		h	MBS	km-MBS	km-MS	GMM-MBS	GMM-MS	Hier-MBS	Hier-MS	$\frac{km-MBS}{MBS}$	$\frac{GMM-MBS}{MBS}$
NC	1	0.08	0.08	0.11	0.07	0.11	0.08	0.11	1.03	0.92	* 1.00
	3	0.19	0.19	0.35	0.02	0.35	0.19	0.35	1.04	0.10	1.00
	6	0.41	0.43	0.74	0.04	0.74	0.41	0.73	1.04	0.10	1.00
	12	0.81	0.80	1.71	0.82	1.72	0.81	1.71	0.98	1.01	1.00
ND	1	0.06	0.06	0.07	0.06	0.07	0.06	0.07	1.00	1.00	1.00
	3	0.15	0.15	0.20	0.15	0.20	0.15	0.20	1.00	1.00	1.00
	6	0.26	0.26	0.45	0.26	0.46	0.26	0.45	0.99	1.00	1.00
	12	0.40	0.39	0.86	0.39	0.88	0.39	0.86	1.00	1.00	1.00
NE	1	0.06	0.06	0.08	0.06	0.08	0.06	0.08	1.01	0.85	* 1.00
	3	0.16	0.16	0.23	0.15	0.23	0.16	0.23	1.00	0.93	* 1.00
	6	0.26	0.26	0.44	0.24	0.44	0.26	0.45	1.01	0.94	* 1.00
	12	0.46	0.46	0.82	0.44	0.82	0.46	0.84	1.01	0.96	* 1.00
NH	1	0.07	0.07	0.07	0.07	0.07	0.07	0.07	1.03	1.03	1.00
	3	0.20	0.20	0.21	0.20	0.21	0.20	0.21	1.02	1.02	1.00
	6	0.35	0.36	0.48	0.44	0.48	0.35	0.48	1.02	1.24	1.00
	12	0.61	0.63	1.08	0.61	1.08	0.60	1.08	1.04	1.01	1.00
NJ	1	0.07	0.07	0.09	0.07	0.09	0.06	0.09	0.91	* 0.91	* 0.88 *
	3	0.19	0.16	0.24	0.16	0.24	0.16	0.24	0.87	* 0.87	* 0.84 *
	6	0.44	0.39	0.56	0.39	0.56	0.38	0.56	0.89	* 0.89	* 0.86 *
	12	0.78	0.79	1.32	0.79	1.32	0.80	1.33	1.02	1.02	1.03
NM	1	0.05	0.05	0.08	0.05	0.08	0.05	0.08	1.00	1.00	1.00
	3	0.13	0.13	0.19	0.13	0.19	0.13	0.19	0.99	1.00	1.00
	6	0.30	0.29	0.38	0.30	0.39	0.30	0.38	0.99	1.00	1.00
	12	0.61	0.60	0.77	0.61	0.86	0.61	0.79	0.99	1.00	1.00
NV	1	0.06	0.06	0.08	0.06	0.08	0.06	0.08	0.93	* 0.93	* 0.97 *
	3	0.16	0.15	0.26	0.15	0.26	0.15	0.26	0.93	* 0.93	* 0.96 *
	6	0.36	0.35	0.63	0.35	0.63	0.35	0.63	0.97	* 0.97	* 0.98 *
	12	0.75	0.77	1.40	0.77	1.41	0.75	1.41	1.02	1.02	1.01
NY	1	0.07	0.07	0.09	0.07	0.09	0.07	0.09	0.92	* 0.92	* 0.90 *
	3	0.19	0.17	0.28	0.17	0.28	0.17	0.28	0.91	* 0.91	* 0.88 *
	6	0.41	0.38	0.59	0.38	0.59	0.37	0.59	0.92	* 0.92	* 0.90 *
	12	0.66	0.64	1.28	0.64	1.29	0.64	1.29	0.97	* 0.97	* 0.97 *

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.13: Multivariate post-sample forecast accuracy for clustered states with
 JD for UR of states OH,OK,OR,PA,RI,SC,SD,TN.

Series	Steps	RMSE						RRMSE				
		h	MBS	km-MBS	km-MS	GMM-MBS	GMM-MS	Hier-MBS	Hier-MS	$\frac{km-MBS}{MBS}$	$\frac{GMM-MBS}{MBS}$	$\frac{Hier-MBS}{MBS}$
OH	1	0.06	0.06	0.09	0.06	0.09	0.06	0.09	1.01	1.00	1.00	
	3	0.16	0.16	0.30	0.16	0.30	0.16	0.30	1.02	1.00	1.00	
	6	0.38	0.38	0.70	0.38	0.70	0.38	0.70	1.01	1.00	1.00	
	12	0.82	0.81	1.60	0.82	1.60	0.82	1.59	0.99	1.00	1.00	
OK	1	0.08	0.08	0.10	0.08	0.10	0.08	0.10	0.97	* 1.00	1.00	
	3	0.21	0.21	0.30	0.21	0.31	0.21	0.30	0.97	* 1.00	1.00	
	6	0.40	0.39	0.72	0.40	0.72	0.40	0.72	0.97	* 1.00	1.00	
	12	0.72	0.69	1.58	0.72	1.60	0.72	1.59	0.96	* 1.00	1.00	
OR	1	0.07	0.07	0.11	0.07	0.11	0.08	0.11	1.02	1.02	1.09	
	3	0.20	0.21	0.40	0.21	0.40	0.22	0.40	1.02	1.02	1.09	
	6	0.49	0.49	0.94	0.49	0.95	0.51	0.95	1.01	1.01	1.04	
	12	1.11	1.09	2.19	1.09	2.21	1.01	2.20	0.98	0.98	0.91	
PA	1	0.07	0.07	0.09	0.07	0.09	0.07	0.09	1.02	1.00	1.00	
	3	0.15	0.16	0.25	0.15	0.25	0.15	0.25	1.03	1.00	1.00	
	6	0.31	0.32	0.48	0.31	0.48	0.31	0.48	1.03	1.00	1.00	
	12	0.54	0.54	1.02	0.54	1.03	0.54	1.02	1.00	1.00	1.00	
RI	1	0.06	0.06	0.08	0.06	0.08	0.06	0.08	1.04	1.08	1.00	
	3	0.14	0.15	0.23	0.14	0.23	0.14	0.23	1.05	1.00	1.00	
	6	0.32	0.34	0.48	0.31	0.48	0.32	0.48	1.08	0.98	1.00	
	12	0.70	0.73	0.95	0.65	0.95	0.69	0.94	1.04	0.93	* 1.00	
SC	1	0.07	0.08	0.11	0.07	0.11	0.07	0.11	1.03	1.00	1.00	
	3	0.18	0.19	0.38	0.18	0.38	0.18	0.38	1.03	1.00	1.00	
	6	0.39	0.40	0.83	0.39	0.83	0.39	0.83	1.03	1.00	1.00	
	12	0.88	0.85	2.02	0.87	2.04	0.88	2.01	0.97	* 1.00	1.00	
SD	1	0.07	0.07	0.12	0.06	0.12	0.07	0.12	1.00	0.90	* 1.00	
	3	0.16	0.16	0.31	0.15	0.31	0.16	0.31	1.00	0.92	* 1.00	
	6	0.29	0.29	0.67	0.28	0.67	0.29	0.67	1.00	0.97	* 1.00	
	12	0.54	0.54	1.38	0.52	1.36	0.54	1.36	1.00	0.96	* 1.00	
TN	1	0.07	0.07	0.11	0.07	0.11	0.07	0.11	1.00	1.00	1.00	
	3	0.20	0.20	0.36	0.20	0.36	0.20	0.36	1.00	1.00	1.00	
	6	0.46	0.46	0.90	0.46	0.90	0.46	0.90	1.00	1.00	1.00	
	12	0.97	0.97	2.15	0.97	2.03	0.97	2.03	1.00	1.00	1.00	

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.14: Multivariate post-sample forecast accuracy for clustered states with
 JD for UR of states TX,UT,VA,VT,WA,WI,WV,WY.

Series	Steps	RMSE						RRMSE				
		h	MBS	km-MBS	km-MS	GMM-MBS	GMM-MS	Hier-MBS	Hier-MS	$\frac{km-MBS}{MBS}$	$\frac{GMM-MBS}{MBS}$	$\frac{Hier-MBS}{MBS}$
TX	1	0.06	0.06	0.07	0.06	0.07	0.06	0.07	1.00	1.00	1.00	
	3	0.17	0.17	0.21	0.17	0.21	0.17	0.21	1.00	1.00	1.00	
	6	0.36	0.36	0.47	0.36	0.47	0.36	0.47	1.00	1.00	1.00	
	12	0.65	0.65	1.05	0.65	1.05	0.65	1.04	1.00	1.00	1.00	
UT	1	0.14	0.14	0.22	0.14	0.22	0.14	0.22	1.01	1.01	1.00	
	3	0.25	0.25	0.63	0.22	0.63	0.25	0.63	1.01	0.88	1.00	
	6	0.54	0.55	1.25	0.53	1.25	0.54	1.24	1.00	0.98	1.00	
	12	0.98	0.98	2.74	0.97	2.73	0.98	2.66	1.00	0.99	1.00	
VA	1	0.07	0.07	0.09	0.06	0.09	0.07	0.09	1.01	0.86	* 1.00	
	3	0.18	0.18	0.28	0.16	0.28	0.18	0.28	1.01	0.91	* 1.00	
	6	0.35	0.36	0.60	0.34	0.60	0.35	0.60	1.01	0.96	* 1.00	
	12	0.59	0.59	1.30	0.58	1.30	0.59	1.29	1.00	0.98	* 1.00	
VT	1	0.08	0.08	0.10	0.08	0.10	0.08	0.10	1.00	1.00	1.00	
	3	0.20	0.20	0.34	0.20	0.34	0.20	0.34	1.00	1.00	1.00	
	6	0.35	0.35	0.75	0.35	0.75	0.35	0.75	1.00	1.00	1.00	
	12	0.62	0.62	1.58	0.62	1.41	0.62	1.41	1.00	1.00	1.00	
WA	1	0.06	0.06	0.10	0.06	0.10	0.06	0.10	1.00	1.00	1.00	
	3	0.16	0.16	0.32	0.16	0.32	0.16	0.32	1.00	1.00	1.00	
	6	0.39	0.39	0.75	0.39	0.74	0.39	0.74	1.00	1.00	1.00	
	12	0.79	0.79	1.88	0.79	1.67	0.79	1.65	1.00	1.00	1.00	
WI	1	0.06	0.06	0.09	0.06	0.09	0.06	0.09	1.01	0.99	1.00	
	3	0.18	0.18	0.31	0.17	0.31	0.18	0.31	1.01	0.96	* 1.00	
	6	0.40	0.40	0.78	0.38	0.78	0.40	0.78	1.01	0.95	* 1.00	
	12	0.83	0.81	1.88	0.80	1.89	0.83	1.86	0.98	0.96	* 1.00	
WV	1	0.07	0.08	0.10	0.08	0.10	0.11	0.10	1.19	1.19	1.59	
	3	0.18	0.23	0.35	0.23	0.35	0.31	0.35	1.29	1.29	1.76	
	6	0.34	0.41	0.79	0.41	0.79	0.59	0.79	1.21	1.21	1.72	
	12	0.63	0.78	1.60	0.78	1.61	1.09	1.61	1.24	1.23	1.73	
WY	1	0.06	0.06	0.07	0.06	0.07	0.06	0.07	1.00	1.00	1.00	
	3	0.17	0.17	0.23	0.17	0.23	0.17	0.22	1.00	1.00	1.00	
	6	0.35	0.35	0.55	0.35	0.54	0.35	0.54	1.00	1.00	1.00	
	12	0.67	0.67	1.52	0.67	1.34	0.67	1.32	1.00	1.00	1.00	

Note:* indicates results are statistically significant at $\alpha=0.01$.

TABLE B.15: Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states AL,AR,AZ,CA,CO,CT,DE,FL.

Series	Steps	RMSE								RRMSE		
		h	MS	MBS	JD-MS	JD-MBS	Tensor-MS	Tensor-MBS	SOEM-MS	SOEM-MBS	$\frac{JD-MBS}{MBS}$	$\frac{Tensor-MBS}{MBS}$
AL	1	0.13	0.07	0.12	0.12	0.13	0.08	1.23	0.06	1.74	1.10	0.92 *
	3	0.41	0.20	0.50	0.41	0.42	0.21	1.23	0.18	2.05	1.07	0.93 *
	6	0.94	0.45	0.93	0.84	0.96	0.47	1.27	0.44	1.86	1.04	0.98
	12	2.15	1.08	2.10	1.56	2.26	1.00	1.62	1.13	1.44	0.92	1.04
AR	1	0.07	0.06	0.07	0.07	0.07	0.06	0.07	0.05	1.21	1.13	0.94 *
	3	0.20	0.14	0.20	0.19	0.20	0.16	0.20	0.13	1.40	1.18	0.93 *
	6	0.50	0.27	0.50	0.45	0.50	0.31	0.51	0.26	1.66	1.14	0.95 *
	12	1.15	0.44	1.19	0.77	1.15	0.50	1.18	0.44	1.76	1.15	1.00
AZ	1	0.09	0.07	0.09	0.09	0.09	0.07	0.09	0.06	1.29	0.97 *	0.85 *
	3	0.27	0.18	0.29	0.27	0.27	0.18	0.27	0.15	1.46	0.97 *	0.81 *
	6	0.57	0.37	0.60	0.55	0.56	0.36	0.57	0.33	1.47	0.97 *	0.89 *
	12	1.34	0.74	1.37	1.19	1.32	0.72	1.34	0.70	1.61	0.97 *	0.95 *
CA	1	0.10	0.06	0.09	0.09	0.10	0.07	0.10	0.05	1.52	1.05	0.78 *
	3	0.29	0.16	0.33	0.29	0.29	0.16	0.29	0.11	1.84	1.05	0.73 *
	6	0.57	0.36	0.58	0.57	0.58	0.37	0.57	0.30	1.60	1.05	0.85 *
	12	1.27	0.82	1.28	1.19	1.30	0.83	1.25	0.83	1.46	1.01	1.02
CO	1	0.10	0.07	0.09	0.09	0.10	0.07	0.10	0.06	1.22	0.98 *	0.78 *
	3	0.33	0.21	0.31	0.30	0.33	0.21	0.33	0.16	1.44	0.98 *	0.74 *
	6	0.78	0.45	0.77	0.73	0.79	0.44	0.78	0.39	1.62	0.98 *	0.87 *
	12	1.60	0.78	1.59	1.20	1.64	0.79	1.58	0.87	1.54	1.00	1.12
CT	1	0.07	0.06	0.07	0.07	0.07	0.06	0.07	0.05	1.06	0.87 *	0.78 *
	3	0.22	0.18	0.22	0.22	0.22	0.15	0.22	0.12	1.22	0.83 *	0.69 *
	6	0.50	0.38	0.50	0.46	0.50	0.31	0.49	0.26	1.20	0.81 *	0.69 *
	12	1.04	0.75	1.07	0.86	1.04	0.68	1.02	0.62	1.15	0.91 *	0.83 *
DE	1	0.10	0.08	0.10	0.10	0.10	0.08	0.10	0.07	1.29	1.02	0.92 *
	3	0.28	0.17	0.28	0.28	0.28	0.18	0.28	0.16	1.59	1.04	0.89 *
	6	0.58	0.34	0.59	0.55	0.58	0.36	0.58	0.29	1.61	1.04	0.85 *
	12	1.26	0.57	1.29	1.03	1.25	0.57	1.23	0.57	1.81	1.01	1.00
FL	1	0.15	0.09	0.15	0.13	0.16	0.09	0.15	0.07	1.38	1.00	0.79 *
	3	0.39	0.21	0.40	0.38	0.43	0.21	0.39	0.17	1.79	1.00	0.82 *
	6	0.76	0.43	0.79	0.69	0.96	0.43	0.76	0.36	1.62	1.00	0.84 *
	12	1.61	0.81	1.66	1.32	2.86	0.81	1.62	0.81	1.62	1.00	1.00

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.16: Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states GA,IA,ID,IL,IN,KS,KY,LA.

Series	Steps	RMSE								RRMSE		
		h	MS	MBS	JD-MS	JD-MBS	Tensor-MS	Tensor-MBS	SOEM-MS	SOEM-MBS	$\frac{JD-MBS}{MBS}$	$\frac{Tensor-MBS}{MBS}$
GA	1	0.10	0.08	0.10	0.10	0.11	0.08	0.10	0.07	1.35	1.00	0.87 *
	3	0.32	0.19	0.32	0.32	0.32	0.19	0.32	0.16	1.68	1.00	0.85 *
	6	0.59	0.37	0.60	0.57	0.60	0.37	0.59	0.32	1.55	1.00	0.88 *
	12	1.31	0.74	1.32	1.14	1.33	0.73	1.29	0.72	1.54	1.00	0.97 *
IA	1	0.23	0.16	0.23	0.21	0.26	0.16	0.23	0.16	1.34	1.00	0.99
	3	0.54	0.28	0.57	0.35	0.81	0.28	0.55	0.27	1.26	1.00	0.99 *
	6	0.92	0.39	0.98	0.49	2.67	0.39	0.96	0.38	1.25	1.00	0.97 *
	12	1.85	0.68	1.98	0.80	0.81	0.68	2.04	0.68	1.18	1.00	0.99
ID	1	0.08	0.06	0.08	0.08	0.08	0.06	2.59	0.06	1.26	1.01	0.96 *
	3	0.24	0.14	0.24	0.24	0.24	0.14	2.63	0.13	1.71	1.02	0.93 *
	6	0.43	0.29	0.44	0.42	0.43	0.29	2.68	0.27	1.48	1.02	0.96 *
	12	0.84	0.60	0.97	0.80	0.85	0.61	2.97	0.57	1.34	1.03	0.95 *
IL	1	0.12	0.07	0.11	0.11	0.12	0.07	0.12	0.06	1.69	1.01	0.93 *
	3	0.38	0.20	0.38	0.38	0.39	0.20	0.38	0.18	1.95	1.01	0.92 *
	6	0.81	0.39	0.81	0.73	0.82	0.39	0.81	0.37	1.86	1.01	0.95 *
	12	1.72	0.78	1.75	1.22	1.75	0.78	1.74	0.78	1.56	1.00	1.00
IN	1	0.12	0.08	0.12	0.12	0.12	0.08	0.13	0.07	1.50	1.02	0.90 *
	3	0.40	0.22	0.38	0.35	0.40	0.22	0.40	0.20	1.61	1.01	0.92 *
	6	0.94	0.49	0.92	0.72	0.95	0.50	0.96	0.49	1.46	1.00	1.00
	12	2.12	1.17	2.09	1.37	2.17	1.15	2.21	1.21	1.17	0.98	1.03
KS	1	0.12	0.08	0.12	0.12	0.12	0.08	0.12	0.07	1.50	1.02	0.84 *
	3	0.33	0.19	0.33	0.34	0.35	0.20	0.33	0.16	1.78	1.02	0.83 *
	6	0.69	0.34	0.70	0.58	0.78	0.35	0.71	0.31	1.69	1.02	0.91 *
	12	1.47	0.63	1.48	0.90	2.11	0.65	1.56	0.53	1.44	1.03	0.85 *
KY	1	0.10	0.07	0.10	0.10	0.10	0.07	0.10	0.06	1.41	1.00	0.91 *
	3	0.31	0.17	0.30	0.30	0.31	0.17	0.31	0.16	1.74	1.00	0.93 *
	6	0.73	0.41	0.72	0.63	0.74	0.41	0.74	0.40	1.54	1.00	0.98 *
	12	1.69	0.86	1.65	1.20	1.76	0.86	1.75	0.88	1.39	1.00	1.02
LA	1	0.83	0.44	0.85	0.79	0.79	0.44	1.27	0.42	1.81	1.00	0.95 *
	3	2.38	0.81	2.80	1.26	1.29	0.81	1.32	0.79	1.56	1.00	0.98 *
	6	4.29	1.08	6.27	1.37	1.45	1.08	1.51	1.05	1.27	1.00	0.97 *
	12	9.35	1.66	22.01	1.58	1.63	1.66	2.08	1.46	0.95	1.00	0.88 *

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.17: Multivariate post-sample forecast accuracy for clustered states with
 JD, Tensor, SOEM for UR of states MA,MD,ME,MI,MN,MO,MS,MT.

Series	Steps	RMSE								RRMSE		
		h	MS	MBS	JD-MS	JD-MBS	Tensor-MS	Tensor-MBS	SOEM-MS	SOEM-MBS	$\frac{JD-MBS}{MBS}$	$\frac{Tensor-MBS}{MBS}$
MA	1	0.08	0.07	0.08	0.08	0.08	0.06	0.08	0.05	1.13	0.90 *	0.76 *
	3	0.23	0.21	0.23	0.23	0.23	0.18	0.23	0.13	1.11	0.84 *	0.63 *
	6	0.47	0.42	0.48	0.46	0.47	0.38	0.47	0.32	1.09	0.91 *	0.75 *
	12	1.06	0.70	1.04	0.88	1.06	0.63	1.05	0.62	1.26	0.90 *	0.88 *
MD	1	0.09	0.07	0.09	0.09	0.09	0.07	0.09	0.06	1.25	1.02	0.90 *
	3	0.27	0.18	0.27	0.27	0.27	0.18	0.27	0.16	1.50	1.03	0.91 *
	6	0.58	0.33	0.59	0.50	0.58	0.34	0.57	0.29	1.53	1.04	0.90 *
	12	1.20	0.55	1.21	0.88	1.20	0.57	1.17	0.51	1.61	1.04	0.93 *
ME	1	0.09	0.07	0.08	0.08	0.09	0.07	0.09	0.06	1.20	1.02	0.87 *
	3	0.28	0.18	0.27	0.39	0.28	0.19	0.28	0.15	2.14	1.04	0.83 *
	6	0.61	0.35	0.61	0.59	0.61	0.35	0.60	0.33	1.70	1.01	0.96 *
	12	1.32	0.57	1.35	1.10	1.30	0.55	1.28	0.61	1.95	0.98	1.08
MI	1	0.11	0.07	0.10	0.10	0.11	0.07	0.11	0.06	1.43	1.01	0.93 *
	3	0.39	0.20	0.37	0.37	0.39	0.20	0.39	0.18	1.87	1.01	0.92 *
	6	0.95	0.46	0.92	0.84	0.96	0.46	0.96	0.44	1.84	1.00	0.97 *
	12	2.26	1.25	2.21	1.72	2.31	1.24	2.33	1.23	1.37	0.99	0.98 *
MN	1	0.11	0.06	0.10	0.10	0.11	0.06	0.11	0.06	1.65	1.02	0.91 *
	3	0.33	0.17	0.32	0.34	0.34	0.17	0.33	0.15	2.06	1.01	0.93 *
	6	0.75	0.38	0.74	0.66	0.81	0.38	0.77	0.37	1.71	1.00	0.97 *
	12	1.59	0.78	1.59	1.03	1.90	0.77	1.69	0.80	1.32	0.98	1.02
MO	1	0.16	0.08	0.16	0.16	0.16	0.08	0.16	0.07	2.18	1.00	0.94 *
	3	0.46	0.22	0.45	0.46	0.49	0.22	0.47	0.22	2.05	1.00	0.96 *
	6	0.93	0.46	0.94	0.75	1.08	0.46	0.96	0.45	1.65	1.00	0.99
	12	1.93	0.90	1.94	1.22	2.89	0.90	2.08	0.93	1.36	1.00	1.04
MS	1	0.38	0.17	0.41	0.38	0.44	0.19	0.38	0.16	2.26	1.14	0.93 *
	3	0.99	0.48	1.18	0.69	1.45	0.51	0.95	0.46	1.42	1.05	0.96 *
	6	1.98	0.76	2.85	0.97	5.07	0.80	1.83	0.76	1.27	1.05	0.99
	12	4.34	1.09	10.19	1.32	53.08	1.13	3.56	1.11	1.22	1.04	1.02
MT	1	0.06	0.04	0.06	0.06	0.06	0.04	0.06	0.04	1.40	1.00	0.97 *
	3	0.14	0.09	0.14	0.14	0.14	0.09	0.14	0.08	1.64	1.01	0.95 *
	6	0.28	0.20	0.30	0.25	0.28	0.20	0.28	0.18	1.24	1.01	0.93 *
	12	0.55	0.44	0.59	0.60	0.55	0.44	0.54	0.41	1.37	1.01	0.94 *

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.18: Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states NC,ND,NE,NH,NJ,NM,NV,NY.

Series	Steps	RMSE								RRMSE		
		h	MS	MBS	JD-MS	JD-MBS	Tensor-MS	Tensor-MBS	SOEM-MS	SOEM-MBS	$\frac{JD-MBS}{MBS}$	$\frac{Tensor-MBS}{MBS}$
NC	1	0.11	0.08	0.10	0.10	0.11	0.08	0.11	0.06	1.34	1.03	0.82 *
	3	0.35	0.19	0.34	0.33	0.35	0.19	0.35	0.15	1.75	1.04	0.81 *
	6	0.73	0.41	0.73	0.67	0.73	0.43	0.74	0.39	1.62	1.04	0.94 *
	12	1.69	0.81	1.71	1.27	1.71	0.80	1.74	0.91	1.56	0.98	1.11
ND	1	0.07	0.06	0.07	0.07	0.07	0.06	2.58	0.06	1.11	1.00	0.94 *
	3	0.20	0.15	0.20	0.20	0.20	0.15	2.63	0.14	1.34	1.00	0.93 *
	6	0.45	0.26	0.46	0.32	0.45	0.26	2.71	0.23	1.26	0.99	0.91 *
	12	0.85	0.40	0.90	0.45	0.85	0.39	2.98	0.37	1.13	1.00	0.94 *
NE	1	0.08	0.06	0.07	0.07	0.08	0.06	0.08	0.06	1.14	1.01	0.93 *
	3	0.23	0.16	0.23	0.23	0.23	0.16	0.23	0.15	1.41	1.00	0.93 *
	6	0.45	0.26	0.47	0.34	0.45	0.26	0.45	0.23	1.34	1.01	0.91 *
	12	0.83	0.46	0.95	0.55	0.83	0.46	0.83	0.42	1.20	1.01	0.91 *
NH	1	0.07	0.07	0.07	0.07	0.07	0.07	0.52	0.05	1.01	1.03	0.78 *
	3	0.21	0.20	0.21	0.21	0.21	0.20	0.67	0.15	1.05	1.02	0.74 *
	6	0.48	0.35	0.48	0.47	0.48	0.36	1.02	0.31	1.33	1.02	0.88 *
	12	1.07	0.61	1.08	0.80	1.08	0.63	1.98	0.51	1.33	1.04	0.85 *
NJ	1	0.09	0.07	0.09	0.09	0.09	0.07	0.09	0.06	1.26	0.91 *	0.80 *
	3	0.24	0.19	0.24	0.24	0.25	0.16	0.24	0.13	1.30	0.87 *	0.69 *
	6	0.56	0.44	0.57	0.54	0.57	0.39	0.56	0.35	1.23	0.89 *	0.79 *
	12	1.35	0.78	1.43	1.04	1.39	0.79	1.33	0.82	1.34	1.02	1.05
NM	1	0.08	0.05	0.08	0.08	0.08	0.05	0.08	0.05	1.41	1.00	0.93 *
	3	0.19	0.13	0.19	0.19	0.19	0.13	0.19	0.12	1.46	0.99	0.88 *
	6	0.38	0.30	0.40	0.40	0.38	0.29	0.38	0.26	1.34	0.99	0.87 *
	12	0.78	0.61	0.89	0.84	0.78	0.60	0.77	0.55	1.38	0.99	0.90 *
NV	1	0.08	0.06	0.08	0.08	0.08	0.06	2.88	0.05	1.31	0.93 *	0.83 *
	3	0.26	0.16	0.25	0.25	0.26	0.15	2.90	0.13	1.62	0.93 *	0.83 *
	6	0.63	0.36	0.63	0.63	0.63	0.35	2.97	0.34	1.74	0.97 *	0.93 *
	12	1.43	0.75	1.43	1.39	1.42	0.77	3.08	0.80	1.85	1.02	1.07
NY	1	0.09	0.07	0.09	0.09	0.09	0.07	0.09	0.06	1.23	0.92 *	0.79 *
	3	0.28	0.19	0.28	0.28	0.28	0.17	0.28	0.14	1.45	0.91 *	0.73 *
	6	0.60	0.41	0.60	0.56	0.60	0.38	0.59	0.33	1.37	0.92 *	0.81 *
	12	1.31	0.66	1.34	1.02	1.34	0.64	1.30	0.68	1.54	0.97	1.03

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.19: Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states OH,OK,OR,PA,RI,SC,SD,TN.

Series	Steps	RMSE								RRMSE		
		h	MS	MBS	JD-MS	JD-MBS	Tensor-MS	Tensor-MBS	SOEM-MS	SOEM-MBS	$\frac{JD-MBS}{MBS}$	$\frac{Tensor-MBS}{MBS}$
OH	1	0.09	0.06	0.09	0.09	0.09	0.06	0.09	0.06	1.44	1.01	0.92 *
	3	0.30	0.16	0.29	0.28	0.30	0.16	0.30	0.14	1.70	1.02	0.87 *
	6	0.69	0.38	0.69	0.60	0.70	0.38	0.70	0.35	1.58	1.01	0.93 *
	12	1.58	0.82	1.57	1.23	1.58	0.81	1.62	0.81	1.50	0.99	0.99
OK	1	0.10	0.08	0.10	0.09	0.10	0.08	2.60	0.07	1.18	0.97 *	0.85 *
	3	0.30	0.21	0.30	0.26	0.31	0.21	2.64	0.17	1.23	0.97 *	0.80 *
	6	0.72	0.40	0.72	0.57	0.72	0.39	2.74	0.35	1.42	0.97 *	0.86 *
	12	1.57	0.72	1.60	0.96	1.57	0.69	3.02	0.67	1.33	0.96 *	0.94 *
OR	1	0.11	0.07	0.10	0.10	0.11	0.07	1.42	0.06	1.46	1.02	0.87 *
	3	0.40	0.20	0.38	0.38	0.40	0.21	1.46	0.18	1.89	1.02	0.88 *
	6	0.96	0.49	0.93	0.78	0.97	0.49	1.68	0.48	1.59	1.01	0.99
	12	2.27	1.11	2.24	1.50	2.34	1.09	2.51	1.23	1.35	0.98	1.10
PA	1	0.09	0.07	0.09	0.09	0.09	0.07	0.09	0.06	1.44	1.02	0.93 *
	3	0.25	0.15	0.25	0.26	0.25	0.16	0.25	0.14	1.69	1.03	0.90 *
	6	0.48	0.31	0.49	0.43	0.48	0.32	0.47	0.29	1.39	1.03	0.95 *
	12	1.01	0.54	1.05	0.78	1.01	0.54	1.00	0.56	1.45	1.00	1.03
RI	1	0.08	0.06	0.08	0.08	0.08	0.06	2.90	0.05	1.37	1.04	0.84 *
	3	0.23	0.14	0.23	0.23	0.23	0.15	2.93	0.11	1.62	1.05	0.81 *
	6	0.48	0.32	0.48	0.48	0.49	0.34	3.02	0.25	1.51	1.08	0.79 *
	12	0.94	0.70	0.93	0.84	0.96	0.73	3.39	0.58	1.21	1.04	0.83 *
SC	1	0.11	0.07	0.10	0.10	0.11	0.08	0.11	0.06	1.40	1.03	0.83 *
	3	0.38	0.18	0.36	0.36	0.38	0.19	0.38	0.16	1.98	1.03	0.88 *
	6	0.82	0.39	0.82	0.74	0.83	0.40	0.84	0.37	1.92	1.03	0.94 *
	12	1.99	0.88	1.99	1.43	2.03	0.85	2.04	0.97	1.63	0.97	1.11
SD	1	0.12	0.07	0.12	0.12	0.12	0.07	0.12	0.07	1.78	1.00	0.96 *
	3	0.31	0.16	0.31	0.31	0.35	0.16	0.32	0.16	1.87	1.00	0.97 *
	6	0.66	0.29	0.67	0.48	0.83	0.29	0.68	0.28	1.67	1.00	0.96 *
	12	1.33	0.54	1.36	0.74	2.54	0.54	1.44	0.52	1.37	1.00	0.96 *
TN	1	0.11	0.07	0.10	0.10	0.11	0.07	0.11	0.07	1.40	1.00	0.92 *
	3	0.36	0.20	0.34	0.31	0.36	0.20	0.36	0.18	1.58	1.00	0.93 *
	6	0.90	0.46	0.88	0.72	0.92	0.46	0.91	0.45	1.57	1.00	0.99
	12	2.01	0.97	1.98	1.39	2.12	0.97	2.07	0.97	1.43	1.00	1.00

Note: * indicates results are statistically significant at $\alpha=0.01$.

TABLE B.20: Multivariate post-sample forecast accuracy for clustered states with JD, Tensor, SOEM for UR of states TX,UT,VA,VT,WA,WI,WV,WY.

Series	Steps	RMSE								RRMSE		
		h	MS	MBS	JD-MS	JD-MBS	Tensor-MS	Tensor-MBS	SOEM-MS	SOEM-MBS	$\frac{JD-MBS}{MBS}$	$\frac{Tensor-MBS}{MBS}$
TX	1	0.07	0.06	0.07	0.07	0.07	0.06	0.07	0.05	1.16	1.00	0.81 *
	3	0.21	0.17	0.20	0.20	0.21	0.17	0.21	0.12	1.19	1.00	0.71 *
	6	0.47	0.36	0.47	0.42	0.47	0.36	0.47	0.27	1.16	1.00	0.76 *
	12	1.04	0.65	1.04	0.77	1.06	0.65	1.06	0.57	1.20	1.00	0.88 *
UT	1	0.22	0.14	0.22	0.22	0.22	0.14	0.99	0.14	1.55	1.01	0.98
	3	0.62	0.25	0.62	0.61	0.72	0.25	1.01	0.24	2.44	1.01	0.96 *
	6	1.22	0.54	1.23	0.99	1.85	0.55	1.04	0.53	1.83	1.00	0.98
	12	2.59	0.98	2.57	1.61	9.10	0.98	1.36	1.01	1.64	1.00	1.03
VA	1	0.09	0.07	0.09	0.09	0.09	0.07	0.09	0.06	1.24	1.01	0.86 *
	3	0.28	0.18	0.27	0.27	0.28	0.18	0.28	0.15	1.54	1.01	0.87 *
	6	0.60	0.35	0.60	0.51	0.61	0.36	0.60	0.34	1.44	1.01	0.96 *
	12	1.28	0.59	1.30	0.88	1.33	0.59	1.25	0.65	1.48	1.00	1.10
VT	1	0.10	0.08	0.09	0.09	0.10	0.08	0.10	0.07	1.22	1.00	0.87 *
	3	0.34	0.20	0.32	0.32	0.34	0.20	0.34	0.17	1.65	1.00	0.85 *
	6	0.75	0.35	0.74	0.58	0.75	0.35	0.74	0.34	1.65	1.00	0.97 *
	12	1.39	0.62	1.37	0.84	1.40	0.62	1.36	0.71	1.35	1.00	1.15
WA	1	0.09	0.06	0.09	0.09	0.10	0.06	0.09	0.06	1.45	1.00	0.93 *
	3	0.32	0.16	0.31	0.31	0.32	0.16	0.32	0.15	1.92	1.00	0.90 *
	6	0.73	0.39	0.73	0.62	0.75	0.39	0.72	0.38	1.61	1.00	0.98 *
	12	1.62	0.79	1.67	1.12	1.70	0.79	1.59	0.83	1.41	1.00	1.05
WI	1	0.09	0.06	0.09	0.08	0.09	0.06	0.09	0.06	1.38	1.01	0.92 *
	3	0.31	0.18	0.30	0.26	0.31	0.18	0.31	0.16	1.48	1.01	0.93 *
	6	0.78	0.40	0.76	0.61	0.79	0.40	0.79	0.39	1.52	1.01	0.98
	12	1.84	0.83	1.80	1.17	1.89	0.81	1.90	0.91	1.41	0.98	1.09 *
WV	1	0.10	0.07	0.10	0.10	0.10	0.08	0.76	0.07	1.38	1.19	0.94 *
	3	0.35	0.18	0.34	0.34	0.35	0.23	0.77	0.16	1.92	1.29	0.93 *
	6	0.79	0.34	0.79	0.69	0.80	0.41	0.86	0.32	2.02	1.21	0.95 *
	12	1.63	0.63	1.68	1.18	1.63	0.78	1.31	0.62	1.88	1.24	0.99
WY	1	0.07	0.06	0.07	0.07	0.18	0.06	0.07	0.05	1.21	1.00	0.90 *
	3	0.22	0.17	0.22	0.22	0.23	0.17	0.22	0.14	1.33	1.00	0.84 *
	6	0.54	0.35	0.54	0.50	0.55	0.35	0.54	0.32	1.42	1.00	0.90 *
	12	1.31	0.67	1.33	0.99	1.23	0.67	1.32	0.61	1.49	1.00	0.91 *

Note: * indicates results are statistically significant at $\alpha=0.01$.

Bibliography

- H. Abdi & L. J. Williams (2010). ‘Principal component analysis’. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(4):433–459.
- M. R. Allen & L. A. Smith (1996). ‘Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise’. *Journal of Climate* **9**(12):3373–3404.
- L. Anselin (1995). ‘Local indicators of spatial association-LISA’. *Geographical analysis* **27**(2):93–115.
- L. Anselin (2013). *Spatial econometrics: methods and models*, vol. 4. Springer Science & Business Media.
- R. O. Awichi & W. G. Müller (2013). ‘Improving SSA predictions by inverse distance weighting’. *REVSTAT–Statistical Journal* **11**(1):105–119.
- S. Aydn, et al. (2011). ‘Singular Spectrum Analysis of Sleep EEG in Insomnia’. *Journal of Medical Systems* **35**:457–461.
- T. Babaie, et al. (2014). ‘Network Traffic Decomposition for Anomaly Detection’. *arXiv preprint arXiv:1403.0157*.
- J. Bai & P. Perron (2003). ‘Computation and analysis of multiple structural change models’. *Journal of Applied Econometrics* **18**(1):1–22.
- L. Ball, et al. (2015). ‘Do forecasters believe in Okuns Law? An assessment of unemployment and output forecasts’. *International Journal of Forecasting* **31**(1):176–184.
- P. C. Bandi, Federico M and Phillips (2003). ‘Fully nonparametric estimation of scalar diffusion models’. *Econometrica* **71**(1):241–283.
- A. Belouchrani, et al. (1997). ‘A blind source separation technique using second-order statistics’. *IEEE Transactions on signal processing* **45**(2):434–444.
- H. J. Bierens (1988). ‘ARMA memory index modeling of economic time series’. *Econometric Theory* **4**(01):35–59.
- B. Blankertz, et al. (2008). ‘Optimizing Spatial filters for Robust EEG Single-Trial Analysis’. *IEEE Signal Processing Magazine* **25**(1):41–56.

- P. Bonizzi, et al. (2014). ‘Singular Spectrum Decomposition: a New Method for Time Series Decomposition’. *Advances in Adaptive Data Analysis* **06**(04):1450011.
- G. E. Box, et al. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- P. J. Brockwell & R. A. Davis (2002). ‘Introduction to Time Series and Forecasting, Second Edition’. p. 434. Springer Texts in Statistics, 2 edn.
- D. S. Broomhead & G. P. King (1986). ‘Extracting qualitative dynamics from experimental data’. *Physica D: Nonlinear Phenomena* **20**(2-3):217–236.
- V. Buchstaber (1994). ‘Time series analysis and grassmannians’. In S. Gindikin, editor, *Applied Problems of Radon Transform* **162**:1–17.
- A. Bunse-Gerstner, et al. (1993). ‘Numerical methods for simultaneous diagonalization’. *SIAM Journal on Matrix Analysis and Applications* **14**(4):927–949.
- L. Cao, et al. (1998). ‘Dynamics from multivariate time series’. *Physica D: Nonlinear Phenomena* **121**:75–88.
- J.-F. Cardoso (1999). ‘High-order contrasts for independent component analysis’. *Neural computation* **11**(1):157–192.
- J.-F. Cardoso & A. Souloumiac (1993). ‘Blind beamforming for non-Gaussian signals’. In *IEE Proceedings F-Radar and Signal Processing*, vol. 140, pp. 362–370. IET.
- J.-F. Cardoso & A. Souloumiac (1996). ‘Jacobi angles for simultaneous diagonalization’. *SIAM journal on matrix analysis and applications* **17**(1):161–164.
- F. Chan (2013). ‘Advantages of Non-Normality in Testing Cointegration Rank’. Available at SSRN 2223753 .
- K. S. Chan & H. Tong (1986). ‘On estimating thresholds in autoregressive models’. *Journal of time series analysis* **7**(3):179–190.
- Y.-W. Chang, et al. (2015). ‘Identification of Basin Topography Characteristic Using Multivariate Singular Spectrum Analysis: Case Study of the Taipei Basin’. *Engineering Geology* **197**:240–252.
- H. S. Chao & C. H. Loh (2014). ‘Application of singular spectrum analysis to structural monitoring and damage diagnosis of bridges’. *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance* **10**:708–727.
- C. Chatfield (2013). *The Analysis of Time Series: An Introduction*. CRC Press, sixth edit edn.
- Q. Chen, et al. (2013). ‘Singular spectrum analysis for modeling seasonal signals from GPS time series’. *Journal of Geodynamics* **72**:25–35.

- P. F. Christoffersen (1998). 'Evaluating interval forecasts'. *International economic review* pp. 841–862.
- A. Cliff & K. Ord (1969). 'The Problem of Spatial Autocorrelation'. In *London Papers in Regional Science 1, Studies in Regional Science* p. 2555.
- A. Cliff & K. Ord (1972). 'Testing for spatial autocorrelation among regression residuals'. *Geographical Analysis* **4**(3):267–284.
- P. Comon (1997). 'Cumulant tensors'. In *HOST 1997, IEEE Signal Processing Workshop on Higher-Order Statistics, July 21-23, 1997, Banff, Canada, Banff, CANADA*.
- I. Daly, et al. (2013). 'Automated Artifact Removal From the Electroencephalogram: A Comparative Study'. *Clinical EEG and Neuroscience* **44**(4):291–306.
- D. Danilov & A. Zhigljavsky (1997). *Principal Components of Time Series: the Caterpillar Method*. St.Petersburg Press, St. Petersburg (in Russian).
- R. a. Davis, et al. (2006). 'Structural Break Estimation for Nonstationary Time Series Models'. *Journal of the American Statistical Association* **101**(473):223–239.
- N. E. Day (1969). 'Estimating the components of a mixture of normal distributions'. *Biometrika* **56**(3):463–474.
- M. L. de Menezes, et al. (2014). 'Combining singular spectrum analysis and PAR(p) structures to model wind speed time series'. *Journal of Systems Science and Complexity* **27**(1):29–46.
- P. Del Moral (1996). 'Nonlinear filtering: Interacting particle solution'. *Markov Processes and Related Fields* **2**(4):555–580.
- A. P. Dempster, et al. (1977). 'Maximum likelihood from incomplete data via the EM algorithm'. *Journal of the royal statistical society. Series B (methodological)* pp. 1–38.
- D. Dickey, et al. (1991). *Time Series: Theory and Methods*, vol. 31.
- J. Durbin & S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, second edi edn.
- M. Elias & S. Rey (2007). *Educational Performance and Spatial Convergence in Peru*. Ph.D. thesis, San Diego State University.
- J. B. Elsner & A. A. Tsonis (1996). *Singular spectrum analysis: a new tool in time series analysis*. Springer, New York.
- P. Erasto & L. Holmstrom (2005). 'Bayesian multiscale smoothing for making inferences about features in scatterplots'. *Journal of Computational and Graphical Statistics* **14**(3):569–589.

- J. Fan & Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.
- D. Fay & D. Rahmani (2017). ‘Clustering time series based on their singular spectrum covariances using a self organising eigenspace map’. In *TBD*. In preparation.
- D. Fay & E. Yoneki (2011). ‘On joint diagonalisation for dynamic network analysis’. *October* (806).
- Y. Feliks, et al. (2010). ‘Oscillatory climate modes in the Eastern Mediterranean and their synchronization with the North Atlantic Oscillation’. *Journal of Climate* **23**(15):4060–4079.
- R. Fendel, et al. (2011). ‘Do professional forecasters believe in the Phillips curve? evidence from the G7 countries’. *Journal of Forecasting* **30**(2):268–287.
- M. R. Fengler, et al. (2003). ‘The dynamics of implied volatilities: A common principal components approach’. *Review of Derivatives Research* **6**(3):179–202.
- C. Floros (2005). ‘Forecasting the UK unemployment rate: model comparisons’. *International Journal of Applied Econometrics and Quantitative Studies* **2**(4):57–72.
- B. N. Flury (1984). ‘Common Principal Components in K Groups’. *Journal of the American Statistical Association* **79**(388):892–898.
- K. Fraedrich (1986a). ‘Estimating dimensions of weather and climate attractors’. *Journal of Atmos Sci* **43**:419–432.
- K. Fraedrich (1986b). ‘Estimating the Dimensions of Weather and Climate Attractors’. *Journal of the Atmospheric Sciences* **43**(5):419–432.
- P. H. Franses (1998). *Time Series Models for Business and Economic Forecasting*. Cambridge University Press.
- P. H. Franses & V. D. Dick (2000). *Non-Linear Time Series Models in Empirical Finance*. Cambridge University Press.
- J. H. Friedman (1987). ‘Exploratory projection pursuit’. *Journal of the American statistical association* **82**(397):249–266.
- J. H. Friedman & W. Stuetzle (1982). ‘Smoothing of scatterplots’. Tech. rep.
- J. H. Friedman & J. W. Tukey (1974). ‘A projection pursuit algorithm for exploratory data analysis’.
- B. Fritzke (1993). ‘Kohonen feature maps and growing cell structuresa performance comparison’. In *Advances in neural information processing systems 5*. Citeseer.

- Gallant, A. Ronald & G. Tauchen. (1989). ‘Semiparametric estimation of conditional constrained heterogeneous processes: Asset pricing application’. *Econometrica: Journal of the Econometric Society* pp. 1091–1120.
- J. Gao (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman and Hall, London.
- J. Gao, et al. (2009). ‘Specification testing in nonlinear and nonstationary time series autoregression’. *Annals of Statistics* **37**(6 B):3893–3928.
- J. Gao & H. Tong (2004). ‘Semiparametric non-linear time series model selection’. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **66**(2):321–336.
- A. Gelb (1974). *Applied optimal estimation*. MIT press.
- D. Gelfond (1967). ‘Finite Difference calculus’ .
- V. Georgescu & S.-M. Delureanu (2015). ‘Analysis using Multivariate and Complex Extensions to Singular Spectrum Analysis’. In ComplexSSA2015 (ed.), *IEEE International Conference*.
- J. Geweke & C. Whiteman (2006). ‘Chapter 1 Bayesian Forecasting’ **1**(05):3–80.
- F. Ghaderi, et al. (2011). ‘Localizing heart sounds in respiratory signals using singular spectrum analysis’. *IEEE Transactions on Biomedical Engineering* **58**(12):3360–3367.
- M. Ghil, et al. (2002). ‘Advanced spectral methods for climatic time series’. *Reviews of geophysics* **40**(1).
- N. Golyandina (2010). ‘On the choice of parameters in singular spectrum analysis and related subspace-based methods’. *arXiv preprint arXiv:1005.4374* .
- N. Golyandina, et al. (2001). *Analysis of Time Series Structure SSA and Related Techniques*. Chapman and Hall, CRC, Boca Raton.
- N. Golyandina & K. D. Usevich (2004). ‘2D-extension of Singular Spectrum Analysis: algorithm and elements of theory’. *Matrix Methods: Theory, Algorithms and Applications* pp. 449–473.
- N. Golyandina & A. Zhigljavsky (2013). *Singular Spectrum Analysis for time series*. Springer Science & Business Media.
- N. E. Golyandina & A. Shlemov (2013). ‘Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series’ (November):1–32.
- C. J. Granger & T. Teräsvirta (1993). *Modelling Nonlinear Dynamic Relationships*. Oxford Univ. Press, Oxford.

- C. W. Granger, et al. (1993). 'Modelling non-linear economic relationships'. *OUP Catalogue*.
- A. Granger, Clive William John and Timmermann (2006). *Handbook of Economic Forecasting*, vol. 1. Elsevier.
- M. S. Grewal (2011). *Kalman filtering*. Springer.
- M. S. Grewal, et al. (2007). *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons.
- A. Groth & M. Ghil (2011). 'Multivariate singular spectrum analysis and the road to phase synchronization'. *Phys. Rev. E* **84**(3):1–12.
- A. Groth, et al. (2011). 'The role of oscillatory modes in US business cycles'. *Journal of Economic Dynamics and Control November* (1).
- R. Gupta & A. Kabundi (2011). 'Forecasting macroeconomic variables using large datasets: Dynamic factor model versus large-scale BVARs'. *Indian Economic Review* pp. 23–40.
- V. Haggan, et al. (1984). 'A study of the application of state-dependent models in non-linear time series analysis'. *Journal of Time Series Analysis* **5**:69–102.
- V. Haggan & T. Ozaki (1981). 'Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model'. *Biometrika* **68**(1):189–196.
- P. Hall (1989). 'On polynomial-based projection indices for exploratory projection pursuit'. *The Annals of Statistics* pp. 589–605.
- J. Hamilton (1994). *Time series analysis*. Princeton Univ. Press, Princeton, NJ.
- W. Hardle, et al. (1996). 'A Review of Nonparametric Time Series Analysis'. *International Statistical Review* **65**(1):49–72.
- L. Hardle, Wolfgang and Tsybakov, Alexandre and Yang (1998). 'Nonparametric vector autoregression'. *Journal of Statistical Planning and Inference* **68**(2):221–245.
- S. Harmeling, et al. (2003). 'Kernel-based nonlinear blind source separation'. *Neural Computation* **15**(5):1089–1124.
- J. Harrison & M. West (1987). 'Practical Bayesian Forecasting'. *The Statistician* pp. 115–125.
- J. P. Harrison & C. F. Stevens (1976). 'Bayesian Forecasting'. *Journal of the Royal Statistical Society. Series B (Methodological)* **38**(3):205–247.
- P. J. Harrison (1967). 'Exponential smoothing and short-term sales forecasting'. *Management science* **13**:821–842.

- A. C. Harvey (1984). 'A unified view of statistical forecasting procedures'. *Journal of Forecasting* **3**(3):245–275.
- H. Hassani (2007). 'Singular spectrum analysis: methodology and comparison'. *Journal of Data Science*. **5**:239–257.
- H. Hassani (2010). 'Singular spectrum analysis based on the minimum variance estimator'. *Nonlinear Analysis: Real World Applications* **11**(3):2065–2077.
- H. Hassani, et al. (2009). 'Forecasting European industrial production with Singular Spectrum Analysis'. *International Journal of Forecasting* **25**:103–118.
- H. Hassani, et al. (2013a). 'Forecasting UK industrial production with multivariate singular spectrum analysis'. *Journal of Forecasting* **32**:395–408.
- H. Hassani, et al. (2012). 'Forecasting UK Industrial Production with Multivariate Singular Spectrum Analysis'. *Journal of Forecasting* **408**(May):1–6.
- H. Hassani & R. Mahmoudvand (2013). 'Multivariate Singular Spectrum Analysis: a General View and New Vector Forecasting Approach'. *International Journal of Energy and Statistics* **01**(01):55–83.
- H. Hassani, et al. (2011a). 'Separability and window length in singular spectrum analysis'. *Comptes Rendus Mathematique* **349**(17-18):987–990.
- H. Hassani, et al. (2013b). 'Predicting inflation dynamics with singular spectrum analysis'. *Journal of the Royal Statistical Society A* **176**:743–760.
- H. Hassani & D. Thomakos (2010). 'A review on singular spectrum analysis for economic and financial time series'. *Statistics and its Interface* **3**(3):377–397.
- H. Hassani, et al. (2015). 'Forecasting U.S. Tourist Arrivals using Optimal Singular Spectrum Analysis'. *Tourism Management* **46**:322–335.
- H. Hassani, et al. (2011b). 'Singular spectrum analysis based on the perturbation theory'. *Nonlinear Analysis: Real World Applications* **12**(5):2752–2766.
- H. Hassani & A. Zhigljavsky (2009). 'Singular spectrum analysis: methodology and application to economics data'. *International Journal of Energy and Statistics* **22**:372–394.
- R. J. Hastie, Trevor J and Tibshirani (1990). *Generalized additive models*.
- S. Heravi, et al. (2004). 'Linear versus neural network forecasts for European industrial production series'. *International Journal of Forecasting* **20**:435–446.
- R. J. Hodrick & E. C. Prescott (1997). 'Postwar US business cycles: an empirical investigation'. *Journal of Money, credit, and Banking* pp. 1–16.

- L. Holmström & I. Launonen (2013). ‘Posterior Singular Spectrum Analysis’. *Statistical Analysis and Data Mining* **6**(5):387–402.
- G. Hori (1999). ‘Joint diagonalization and matrix differential equations’. In *Proc. of NOLTA99, IEICE* pp. 675–678.
- Z. Hou, et al. (2014). ‘Periodicity of Carbon Element Distribution Along Casting Direction in Continuous-Casting Billet by Using Singular Spectrum Analysis’. *Metallurgical and Materials Transactions* **45**:1817–1826.
- N. E. Huang, et al. (1998). ‘The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis’. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **454**(1971):903–995.
- P. J. Huber (1985). ‘Projection pursuit’. *The annals of Statistics* pp. 435–475.
- R. J. Hyndman & Y. Khandakar (2008). ‘Automatic time series forecasting : the forecast package for R Automatic time series forecasting : the forecast package for R’. *Journal Of Statistical Software* **27**(3):1–22.
- A. Hyvärinen & E. Oja (2000). ‘Independent component analysis: algorithms and applications’. *Neural networks* **13**(4):411–430.
- K. Ito & K. Xiong (2000). ‘Gaussian filters for nonlinear filtering problems’. *IEEE Transactions on Automatic Control* **45**(5):910–927.
- S. Jevrejeva, et al. (2006). ‘Nonlinear trends and multiyear cycles in sea level records’. *Journal of Geophysical Research: Oceans* **111**(C9).
- S. C. Johnson (1967). ‘Hierarchical clustering schemes’. *Psychometrika* **32**(3):241–254.
- M. Joho & H. Mathis (2002). ‘Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation’. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002*, pp. 273–277. IEEE.
- M. Joho & K. Rahbar (2002). ‘Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation’. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002*, pp. 403–407. IEEE.
- I. Jolliffe (2002). *Principal component analysis*. Wiley Online Library.
- M. C. Jones & R. Sibson (1987). ‘What is projection pursuit?’. *Journal of the Royal Statistical Society. Series A (General)* pp. 1–37.
- S. Julier & J. Uhlmann (1997). ‘A new extension of the Kalman filter to nonlinear systems’. In *Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defense Sensing, Simulations and Controls*.

- S. J. Julier, et al. (1995). ‘A new approach for filtering nonlinear systems’. In *American Control Conference, Proceedings of the 1995*, vol. 3, pp. 1628–1632. IEEE.
- R. Kalman (1960). ‘A new approach to linear filtering and prediction problems’. *Journal of basic Engineering* **82**(1):35–45.
- H. Kantz & T. Schreiber (2004). ‘Nonlinear Time Series Analysis (Kantz, Schreiber)(Cambridge 2004)(2nd ed).pdf’.
- M. Kapl & W. Mueller (2010). ‘Prediction of steel prices: a comparison between a conventional regression model and MSSA’. *Stat Interface* **3**(3):369–375.
- H. A. Karlsen, et al. (2007). ‘Nonparametric estimation in a nonlinear cointegration type model’. *The Annals of Statistics* **35**(1):252—299.
- D. Karlsen, Hans Arnfinn and Tjostheim (2001). ‘Nonparametric estimation in null recurrent time series’. *Annals of Statistics* pp. 372—416.
- M. A. R. Khan & D. Poskitt (2013). ‘A note on window length selection in Singular Spectrum Analysis.’. *Australian and New Zealand Journal of Statistics* **55**(2):87–108.
- T. G. Kolda & B. W. Bader (2008). ‘Tensor Decompositions and Applications’. *SIAM Review* **51**(3):455–500.
- Z. J. Koles (1991). ‘The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG’. *Electroencephalography and clinical Neurophysiology* **79**(6):440–447.
- D. Kondrashov & M. Ghil (2006). ‘Spatio-temporal filling of missing points in geophysical data sets.’. *Nonlinear Processes in Geophysics* **13**(2):151–159.
- G. Koop (2003). *Bayesian Econometrics*. Chichester edn.
- S. M. Kotkin J (2013). ‘A Map of America’s Future: Where Growth Will Be over the Next Decade.’. *New Geography* .
- S. Kouchaki & S. Sanei (2013). ‘Tensor based singular spectrum analysis for nonstationary source separation’. In *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–5. IEEE.
- S. Kouchaki, et al. (2015). ‘Tensor Based Singular Spectrum Analysis for Automatic Scoring of Sleep EEG’. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **23**(1):1–9.
- A. Lahiani & O. Scaillet (2009). ‘Testing for threshold effect in ARFIMA models: Application to US unemployment rate data’. *International Journal of Forecasting* **25**(2):418–428.

- G. N. Lance & W. T. Williams (1967). ‘A general theory of classificatory sorting strategies II. Clustering systems’. *The computer journal* **10**(3):271–277.
- L. D. Lathauwer (2011). ‘A short introduction to tensor-based methods for factor analysis and blind source separation’. *International Symposium on Image and Signal Processing and Analysis* **1**(2):4–6.
- K. Le Bail, et al. (2014). ‘Quantifying the Correlation Between the MEI and LOD Variations by Decomposing LOD with Singular Spectrum Analysis’. In *Earth on the Edge: Science for a Sustainable Planet*, pp. 473–477. Springer.
- Lee, T.-H., H. White & C. W. J. Granger (1993). ‘Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests.’. *Journal of Econometrics* **56**:269–90.
- J. G. Lewis, Peter AW and Stevens (1991). ‘Nonlinear Modeling of Time Series Using Multivariate Adaptive Regression Splines (MARS)’. *Journal of the American Statistical Association* **86**(416):864–877.
- J. Li & S. X. Chen (2012). ‘Two sample tests for high-dimensional covariance matrices’. *Ann. Statist.* **40**(2):908–940.
- F. Lisi, et al. (1995). ‘Combining Singular-Spectrum Analysis and Neural Networks for Time Series Forecasting.’. *Neural Processing Letters* **2**(4):6–10.
- K. Liu, et al. (2014). ‘Singular spectrum analysis for enhancing the sensitivity in structural damage detection’. *Journal of Sound and Vibration* **333**:392–417.
- S. Lloyd (1982). ‘Least squares quantization in PCM’. *IEEE transactions on information theory* **28**(2):129–137.
- H. Lütkepohl (2005). *New introduction to multiple time series analysis*. Springer Science and Business Media.
- J. MacQueen et al. (1967). ‘Some methods for classification and analysis of multivariate observations’. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297. Oakland, CA, USA.
- R. Mahmoudvand, et al. (2013). ‘On the optimal parameters for reconstruction and forecasting in singular spectrum analysis’. *Communications in Statistics-Simulation and Computation* **42**(4):860–870.
- P. S. Maybeck (1982). *Stochastic models, estimation, and control*, vol. 3. Academic press.
- P. McCullagh (1987). *Tensor methods in statistics*.
- G. McLachlan & D. Peel (2004). *Finite mixture models*. John Wiley & Sons.

- R. J. Meinhold & N. D. Singpurwalla (1983). ‘Understanding the Kalman filter’. *The American Statistician* **37**(2):123–127.
- C. Milas & P. Rothman (2008). ‘Out-of-sample forecasting of unemployment rates with pooled STVECM forecasts’. *International Journal of Forecasting* **24**(1):101–121.
- B. Mizrach (1996). ‘Forecast comparison in L2’. Tech. rep., Working Papers, Department of Economics, Rutgers, The State University of New Jersey.
- R. Mohler (1973). *Bilinear Control Process*. New York: Academic Press.
- L. Molgedey & H. G. Schuster (1994). ‘Separation of a mixture of independent signals using time delayed correlations’. *Physical review letters* **72**(23):3634.
- P. A. Moran (1950). ‘Notes on continuous stochastic phenomena’. *Biometrika* pp. 17–23.
- E. Moreau (2001). ‘A generalization of joint-diagonalization criteria for source separation’. *IEEE Transactions on Signal Processing* **49**(3):530–541.
- S. C. Morton (1989). *Interpretable projection pursuit*. Ph.D. thesis, Citeseer.
- V. Moskvina & A. Zhigljavsky (2003). ‘Change-point detection algorithm based on the singular-spectrum analysis’. *Communication in Statistics: Simulation and Computation* **32**:319–352.
- B. Muruganatham, et al. (2013). ‘Roller element bearing fault diagnosis using singular spectrum analysis’. *Mechanical Systems and Signal Processing* **35**:150–166.
- G. Nelson & A. Rae (2016). ‘An economic geography of the United States: from commutes to megaregions’. *PLoS One* **11**(11).
- M. O’Connor, et al. (2001). ‘The asymmetry of judgemental confidence intervals in time series forecasting’. *International Journal of Forecasting* **17**(4):623–633.
- V. Oropeza & M. Sachchi (2011). ‘Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis’. *Geophysics* **76**:25–32.
- D. R. Osborn, et al. (1999). ‘Seasonal unit roots and forecasts of two-digit European industrial production’. *International Journal of Forecasting* **15**(1):27–47.
- K. Patterson, et al. (2011). ‘Multivariate singular spectrum analysis for forecasting revisions to real-time data’. *Journal of Applied Statistics* **38**(10):2183–2211.
- A. Pepelyshev (2010). ‘Comparison of Recurrent and Vector forecasting’ pp. 457–461.
- M. H. Pesaran & S. M. Potter (1992). ‘Nonlinear Dynamics and Econometrics :’. *Journal of applied econometrics* **7**.
- D.-T. Pham & J.-F. Cardoso (2001). ‘Blind separation of instantaneous mixtures of nonstationary sources’. *IEEE Transactions on Signal Processing* **49**(9):1837–1848.

- A. H. Phan & A. Cichocki (2010). ‘Tensor decompositions for feature extraction and classification of high dimensional datasets’. *Nonlinear theory and its applications, IEICE* **1**(1).
- P. C. B. Phillips (1998). ‘Econometric analysis of Fisher’s equation’. *Cowles Found* **1180**:1299–1325.
- G. Plaut & R. Vautard (1994). ‘Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere’. *Journal of the Atmospheric Sciences* **51**(2):210–236.
- J. Pole, Andy and West, Mike and Harrison (1994). *Applied Bayesian forecasting and time series analysis*. CRC Press.
- M. Priestley (1980). ‘State dependent models: A general approach to non-linear time series analysis’. *Journal of Time Series Analysis* **1**(1):47–71.
- M. Priestley & M. Chao (1972). ‘Non-parametric function fitting’. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 385–392.
- M. B. Priestley (1981). ‘On the fitting of general non-linear time series models’. In *4th International time series meeting, Valencia, Amsterdam, North Holland Pub., 1982*.
- K. Pukenas (2014). ‘Algorithm for the characterization of the cross-correlation structure in multivariate time series’. *Circuits, Systems, and Signal Processing* **33**(4):1289–1297.
- M. Qadrdan, et al. (2013). ‘Probabilistic Wind Power Forecasting using a Single Forecast’. *International Journal of Energy and Statistics* **1**(2):99–111.
- D. Rahmani (2014). ‘A forecasting algorithm for Singular Spectrum Analysis based on bootstrap Linear Recurrent Formula coefficients’. *International Journal of Energy and Statistics* **2**(4):287–299.
- D. Rahmani, et al. (2016). ‘Forecasting time series with structural breaks with Singular Spectrum Analysis, using a general form of recurrent formula’. *arXiv: 605.02188* pp. 1–27.
- A. Rényi (1961). ‘On measures of entropy and information’. In *Fourth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 547–561.
- S. Roth & J. J. Reijmer (2005). ‘Holocene millennial to centennial carbonate cyclicity recorded in slope sediments of the Great Bahama Bank and its climatic implications’. *Sedimentology* **52**(1):161–181.
- D. Ruelle (1980). ‘Strange attractors’. *The Mathematical Intelligencer* **2**(3):126–137.
- C. Sammut & G. I. Webb (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.

- S. Sanei, et al. (2011). ‘An adaptive singular spectrum analysis approach to murmur detection from heart sounds’. *Medical engineering & physics* **33**(3):362–367.
- S. Sarkka (2013). ‘Bayesian Filtering and Smoothing’ p. 254.
- T. Sauer, et al. (1991). ‘Embedology’. *Journal of statistical Physics* **65**(3):579–616.
- N. Schanne, et al. (2010). ‘Regional unemployment forecasts with spatial interdependencies’. *International Journal of Forecasting* **26**(4):908–926.
- L. Scrucca (2005). ‘Clustering multivariate spatial data based on local measures of spatial autocorrelation’. *Quaderni del Dipartimento di Economia, Finanza e Statistica* **20**:1–11.
- J. L. F. Serrano & M. D. Robles Fernandez (2001). ‘Structural Breaks and interest rates forecast: a sequential approach’. Documentos de trabajo del icae, Universidad Complutense de Madrid, Facultad de Ciencias Económicas y Empresariales, Instituto Complutense de Análisis Económico.
- M. Skare & V. Buterin (2015). ‘Modelling and Forecasting Unemployment Non-linear Dynamics Using Spectral Analysis’. *Engineering Economics* **26**(9481):373–383.
- M. Steel (2001). ‘Bayesian Time Series Analysis’. *A Course in Time Series Analysis* pp. 1–12.
- C. Stevens & J. P. Harrison (1971). ‘Bayesian Forecasting Chapter 1’. Warwick University.
- C. Stevens & J. P. Harrison (1973). ‘Bayesian Forecasting chapter 1’. Warwick University.
- J. Sun, et al. (2008). ‘Incremental tensor analysis: Theory and applications’. *ACM Trans. Knowl. Discov. Data* **2**(3):1–37.
- F. Takens (1981). ‘Detecting strange attractors in turbulence’. In *Dynamical systems and turbulence, Warwick 1980*, pp. 366–381. Springer.
- R. Tatoian & L. Hamel (2016). ‘Self-Organizing Map Convergence’. In *Proceedings of the International Conference on Data Mining (DMIN)*, p. 92. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- T. Teljdsvirta (1990). *Power properties of linearity tests for time series. Economics*. University of California, San Diego.
- A. Teixeira, et al. (2005). ‘On the use of clustering and local singular spectrum analysis to remove ocular artifacts from electroencephalograms’. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.* **4**(Mdl):2514–2519.

- T. Teräsvirta, et al. (1993). ‘Power of the neural network linearity test’. *Journal of Time Series Analysis* **14**(2):209–220.
- T. Teräsvirta, et al. (2008). *Modelling nonlinear economic time series*.
- D. D. Thomakos (2008). ‘Optimal Linear Filtering, Smoothing and Trend Extraction for Processes with Unit Roots and Cointegration’ .
- R. Thuraisingham (2013). ‘Use of SSA and MCSSA in the Analysis of Cardiac RR Time Series’. *Journal of Computational Medicine, Hindawi Publishing Corporation* .
- D. Tjøstheim & B. H. Auestad (1994). ‘Nonparametric identification of nonlinear time series: projections’. *Journal of the American Statistical Association* **89**(428):1398–1409.
- H. Tong (1990). ‘Nonlinear time series analysis since 1990: some personal reflections’ pp. 1–13.
- H. Tong & K. S. Lim (1980). ‘Threshold autoregression, limit cycles and cyclical data’. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 245–292.
- B. Trendle (2002). ‘Regional variation in Queensland’s unemployment rate’. *Australian and New Zealand Regional Science Association* .
- L. R. Tucker (1966). ‘Some mathematical notes on three-mode factor analysis’. *Psychometrika* **31**(3):279–311.
- P. Tukey & J. Tukey (1981). ‘Preparation; prechosen sequences of views’. In *Barnett, V., editor, Interpreting Multivariate Data* p. 189–213.
- R. D. Turner (2011). ‘Gaussian Processes for State Space Models and Change Point Detection’. *Learning* .
- K. D. Usevich (2010). ‘On signal and extraneous roots in singular spectrum analysis’. *Statistics and Its Interface* **3**(3):281–295.
- A.-J. van der Veen, et al. (1992). ‘Azimuth and elevation computation in high resolution DOA estimation’. *IEEE Transactions on Signal Processing* **40**(7):1828–1832.
- A.-J. Van der Veen, et al. (1998). ‘Joint angle and delay estimation using shift-invariance techniques’. *IEEE Transactions on Signal Processing* **46**(2):405–418.
- R. Vautard & M. Ghil (1989). ‘Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series’. *Physica*. **35**:395–424.
- R. Vautard, et al. (1992). ‘Singular-spectrum analysis: A toolkit for short, noisy chaotic signals’. *Physica D: Nonlinear Phenomena* **58**(1-4):95–126.

- H. Viljoen & D. G. Nel (2010). ‘Common singular spectrum analysis of several time series’. *Journal of Statistical Planning and Inference* **140**(1):260–267.
- E. A. Wan & R. Van Der Merwe (2000). ‘The unscented Kalman filter for nonlinear estimation’. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pp. 153–158. Ieee.
- P. C. Wang, Qiying and Phillips (2009). ‘Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression’. *Econometric Theory* **25**(03):710—738.
- M. West (2013). ‘Bayesian dynamic modelling’. *Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian FM Smith* pp. 145–166.
- M. West & J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*.
- A. Yeredor (2002). ‘Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation’. *IEEE Transactions on signal processing* **50**(7):1545–1553.
- P. Yiou, et al. (2000). ‘Data-adaptive wavelets and multi-scale singular spectrum analysis.’. *Physica* pp. 254–290.
- D. Young, Peter Colin and Ng, Cho Nam and Lane, Kevin and Parker (1991). ‘Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data’. *Journal of Forecasting* **10**(1-2):57—89.
- G. Zhang, et al. (1998). ‘Forecasting with artificial neural networks : The state of the art.’. *International Journal of Forecasting* **14**:35–62.
- Y. Zhang & X. F. Hui (2012). ‘Research on daily exchange rate forecasting with multivariate singular spectrum analysis’. *International Conference on Management Science and Engineering - Annual Conference Proceedings* pp. 1365–1370.
- X. Zhao, et al. (2011). ‘Multifractal Detrended Cross-Correlation Analysis of Chinese Stock Markets Based on Time Delay’. *Fractals* **19**(03):329–338.
- A. A. Zhigljavsky (2010). ‘Singular Spectrum Analysis for time series: Introduction to this special issue’. *Statistics and its Interface* **3**:255–258.
- W.-X. Zhong (2006). *Duality System in Applied Mechanics and Optimal Control*. Springer Science & Business Media.
- A. Ziehe, et al. (2004). ‘A Fast Algorithm for Joint Diagonalization with Non-orthogonal Transformations and its Application to Blind Source Separation’. *Journal of Machine Learning Research* **5**:777–800.
- E. Zivot & J. Wang (2006). ‘Nonlinear Time Series Models’. *Modeling Financial Time Series with S-PLUS®* pp. 651–709.