

# TOPICS IN NONSTATIONARY TIME SERIES ANALYSIS

A Dissertation

by

RAANJU RAGAVENDAR SUNDARARAJAN

Submitted to the Office of Graduate and Professional Studies of  
Texas A&M University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mohsen Pourahmadi
Co-Chair of Committee,	Suhasini Subba Rao
Committee Members,	Xianyang Zhang
	Natarajan Sivakumar
Head of Department,	Valen E. Johnson

August 2018

Major Subject: Statistics

Copyright 2018 Raanju Ragavendar Sundararajan

## ABSTRACT

Several interesting applications in areas such as neuroscience, economics, finance and seismology have led to the collection nonstationary time series data wherein the statistical properties of the observed process change across time. The analysis of nonstationary time series data is an important and challenging task with useful applications. In comparison to stationarity, modeling temporal dependence in nonstationary time series is more non-trivial, and numerous methods have been proposed to tackle this problem. Stationarity in time series is more coveted than nonstationarity and many of the existing techniques attempt to transform the problem of nonstationarity to a stationary time series setting.

Change point detection is one such method that attempts to find time points wherein the statistical properties of the time series changed. We develop a nonparametric method to detect multiple change points in multivariate piecewise stationary processes when the locations and number of change points are unknown. Based on a test statistic that measures differences in the spectral density matrices through the  $L_2$  norm, we sequentially identify points of local maxima in the test statistic and test for the significance of each of them being change points. In addition, the components responsible for the change in the covariance structure at each detected change point are identified. The asymptotic properties of the test for significant change points under the null and alternative hypothesis are derived.

Another related method for handling nonstationarity is the recent technique of stationary subspace analysis (SSA) that aims at finding linear transformations of nonstationary processes that are stationary. We propose an SSA procedure for general multivariate second-order nonstationary processes. It relies on the asymptotic uncorrelatedness of the discrete Fourier transform of a stationary time series to define a measure of departure from stationarity; it is then minimized to find the stationary subspace. The dimension of the subspace is estimated using a sequential testing procedure and its asymptotic properties are discussed. We illustrate the broader applicability and

better performance of our method in comparison to existing SSA methods through simulations and discuss an application in neuroeconomics. Here we apply our method to filter out noise in EEG brain signals from an economic choice task experiment. This improves prediction performance and more importantly reduces the number of trials needed from individuals in neuroeconomic experiments thereby aligning with the principle of simple and controlled designs in experimental and behavioral economics.

## DEDICATION

To my family and friends.

## ACKNOWLEDGMENTS

My first and foremost thanks is towards my advisor Dr. Mohsen Pourahmadi who has taught me so much over the past few years. His superior expertise, vision, commitment and patience have all contributed immensely towards my evolution as an academic. Dr. Pourahmadi was always available for discussions and his unwavering support and encouragement was largely important in getting myself over the finish line. I am delighted to have learned from him and am deeply indebted for his support and guidance.

My sincere thanks to my co-advisor Dr. Suhasini Subba Rao for her continuous guidance and support. Her immense knowledge and expertise in time series has helped me develop a deep interest for this area of statistics. It is fair to say that I am incredibly lucky to have had the opportunity to learn from not one but two eminent scholars.

I would like to extend my gratitude towards Dr. Sivakumar Natarajan for his positive influence since 2010. Many thanks to Dr. Michael Longnecker who gave me the opportunity to teach and I greatly appreciate his help on numerous occasions. Thank you to Dr. Marco Palma who has been an absolute joy to work with. I would like to extend my appreciation towards Julie Carroll who provided some much needed support during my early teaching days. Many thanks to Dr. Xianyang Zhang and Dr. Fred Dahm at the Statistics Department.

Thank you to my mother and my family for the constant support and care. Finally, this would not be possible if not for my friends who stood by me during some very difficult times over the past few years.

## CONTRIBUTORS AND FUNDING SOURCES

### **Contributors**

This work was supervised by a dissertation committee consisting of Professor Mohsen Pourahmadi (Chair), Professor Suhasini Subba Rao (Co-Chair), Professor Xianyang Zhang (Committee Member) of the Department of Statistics and Professor Sivakumar Natarajan of the Department of Mathematics.

Primary contributions in Chapters 2 and 3 was made independently by the student under the supervision of Professor Mohsen Pourahmadi.

The data analyzed for Chapter 3, Section 3.4 was from the Texas A & M Human Behavior Lab led by Professor Marco A. Palma of the Department of Agricultural Economics. Part of the analysis and inference given in Chapter 3, Section 3.4 was provided by Professor Marco A. Palma.

### **Funding Sources**

Graduate study was supported by fellowships from Texas A&M University and research fellowships from National Science Foundation DMS-1513647, National Science Foundation DMS-1612984.

# TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
DEDICATION .....	iv
ACKNOWLEDGMENTS .....	v
CONTRIBUTORS AND FUNDING SOURCES .....	vi
TABLE OF CONTENTS .....	vii
LIST OF FIGURES .....	ix
LIST OF TABLES .....	xi
1. INTRODUCTION .....	1
1.1 Change Point Detection .....	2
1.2 Stationary Subspace Analysis (SSA) .....	4
2. CHANGE POINT DETECTION IN PIECEWISE STATIONARY TIME SERIES .....	8
2.1 Working Model and the Approach in Preuss et al. (2015) .....	8
2.2 An Alternative Measure of Local Change .....	10
2.2.1 Estimating the Measure of Local Change .....	10
2.3 Finding Change Point Locations .....	11
2.3.1 Choice of N: Length of Local Neighborhood .....	12
2.4 Testing for Significant Change Points .....	13
2.4.1 Block Bootstrap for Stationary Processes .....	14
2.5 Identifying the Components Responsible for the Change .....	15
2.6 Constructing Confidence Intervals for Change Points .....	15
2.7 Theoretical Properties .....	16
2.8 Illustrative Examples .....	20
2.8.1 Simulation Study .....	20
2.8.1.1 Example 1: Change Point Detection Accuracy .....	20
2.8.1.2 Example 2: Empirical Size of the Test .....	30
2.8.1.3 Example 3: Impact of Dimension .....	31
2.8.1.4 Example 4: Confidence Intervals for Change Points .....	34
2.8.2 Application to Seismic Data .....	35

3. STATIONARY SUBSPACE ANALYSIS OF NONSTATIONARY PROCESSES .....	39
3.1 The SSA Setup .....	39
3.2 Dependent-SSA (DSSA) .....	41
3.2.1 DSSA for Constant Mean Processes .....	42
3.2.2 DSSA for Time-Varying Mean .....	43
3.2.3 Identifiability of the Stationary Subspace .....	44
3.2.4 Testing for Stationarity .....	46
3.2.5 Estimation of $d$ : Dimension of the Stationary Subspace .....	47
3.2.6 Asymptotic Properties of the Estimated $d$ .....	47
3.3 Simulation Study .....	49
3.4 A Case Study in Neuroeconomics .....	57
3.4.1 Data Acquisition .....	58
3.4.2 Data Analysis .....	59
3.4.2.1 Noise Reduction via SSA .....	60
3.4.3 The Prediction Models .....	63
3.4.3.1 Prediction Performance .....	65
4. SUMMARY AND CONCLUSIONS .....	75
REFERENCES .....	77
APPENDIX A. TECHNICAL PROOFS .....	84
APPENDIX B. ....	91
B.1 Optimization Details from Section 3.2.1 .....	91
B.2 Details from Section 3.2.3 .....	93



## LIST OF FIGURES

FIGURE	Page
2.1 Change point at $\{0.5\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.18) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014). . . . .	24
2.2 Change points at $\{0.3, 0.75\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.19) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014). . . . .	24
2.3 Change points at $\{0.5, 0.75\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.20) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014). . . . .	25
2.4 Change points at $\{0.5, 0.75\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.21) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014). . . . .	25
2.5 Change points at $\{0.5, 0.75\}$ . (a),(c) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (3.20) using Preuss et al. (2015) and (b),(d) is the same histogram using our method at 1% level. . . . .	26
2.6 Change points at $\{0.25, 0.5, 0.75\}$ . (a),(c) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.23) using Preuss et al. (2015) and (b),(d) is the same histogram using our method at 1% level. . . . .	26
2.7 (a),(c) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.24) using Preuss et al. (2015) and (b),(d) is the same histogram using our method at 1% level. . . . .	27
2.8 $T = 400$ observations of the Acceleration(m/s/s) readings during the evening of Jun 10, 1975 at Palmerston North Telephone Exchange. Measurements are presented in three orthogonal directions. Dashed lines are change points detected by our method. . . . .	37

2.9	Components contribution plot: $ M_{\hat{\delta},ij} $ , $i, j = 1, 2, 3$ (y-axis) from (2.8) for time points (x-axis) 51,52,...,350. ....	38
3.1	Band-Pass filtered 9-dimensional EEG signal $\{X_{t,j} : t = 1, 2, \dots, 640\}$ (before noise reduction) gathered from subject # 31 while responding to food-choice question number 9. ....	61
3.2	3-dimensional stationary subspace process $\{Y_{t,j} : t = 1, 2, \dots, 640\}$ (after noise reduction) gathered from subject #31 while responding to food-choice question number 9. ....	62
3.3	Overall prediction accuracy rate ( $A_1$ ), in %, based on a 10-fold cross-validation for the 10 food choice tasks for the two models $M_1$ and $M_2$ . Approximate 95% confidence intervals included for each accuracy estimate. ....	68
3.4	Prediction accuracy rate as average of sensitivity and specificity ( $A_2$ ), in %, based on a 10-fold cross-validation for the 10 food choice questions for the two models $M_1$ and $M_2$ . Approximate 95% confidence intervals included for each accuracy estimate. ....	69
3.5	Average 10-fold cross-validation overall accuracy rate (in %) for the 10 food choice questions (y-axis) versus dimension of the stationary subspace (x-axis). Approximate 95% confidence intervals included for each accuracy estimate. ....	71
3.6	Prediction accuracy rate: average of sensitivity and specificity ( $A_2$ ), in %, based on a 10-fold cross-validation for the 10 food choice questions. Model $M_3$ was used with $d = 8$ . Approximate 95% confidence intervals included for each accuracy estimate. ....	72
3.7	Cross-validation estimate of the AUC in % (Area under the ROC curve) for the 3 models $M_1$ $M_2$ and $M_3$ . Approximate 95% confidence intervals included for each accuracy estimate. For model $M_3$ we take $d = 8$ . ....	73

## LIST OF TABLES

TABLE	Page	
2.1	The mean number of change points detected per simulation run of the 7 models in Section 2.8.1.1 for the 3 competing methods M1 (Preuss et al. (2015)), M2 (proposed method) and M3 (Matteson and James (2014)). $K$ denotes the actual number of change points. The % Correct column lists the percentage of runs the correct number or change points was reported by the methods. Results are based on 500 simulation runs of all 7 models. ....	29
2.2	The empirical size of the proposed procedure based on two norm choices ( $L_2$ and $\infty$ ) used in the test statistic in (2.4). Results are based on 500 simulation runs of the 3 models in Section 2.8.1.2. The true level of significance (size) was fixed at 1%. ....	31
2.3	(Based on $L_2$ norm) The size (first column) and the empirical power (body of the table) of the proposed test against dimension ( $p$ ) of the input series from 500 simulation runs of the two models given in Section 2.8.1.3. ....	32
2.4	(Based on $\infty$ -norm) The size (first column) and the empirical power (body of the table) of the proposed test against dimension ( $p$ ) of the input series from 500 simulation runs of the two models given in Section 2.8.1.3. ....	33
2.5	Mean and Median of $\bar{b}$ from (2.10) over 500 runs. Mean 90% CI limits and estimated coverage probability is listed for the 7 models (a)-(g) from Section 2.8.1.1 ....	35
3.1	Study 1: $p = 3$ , Independent Sources: Average p-values of test of stationarity for $d=1, 2, 3$ for the three competing methods. True values of $d$ are given in second column. ....	51
3.2	Study 1: $p = 5$ , Independent Sources: Average p-values of test of stationarity for $d=1, 2, 3, 4, 5$ for the three competing methods. True values of $d$ are given in second column. ....	52
3.3	Study 1: $p = 3$ , Dependent Sources: Average p-values of test of stationarity for $d=1, 2, 3$ for the three competing methods. True values of $d$ are given in second column. ....	53

3.4	Study 1: $p = 5$ , Dependent Sources: Average p-values of test of stationarity for $d=1, 2, 3, 4, 5$ for the three competing methods. True values of $d$ are given in second column. ....	54
3.5	Study 2: Number of times the true dimension $d$ was detected by DSSA, ISSA and ASSA for the models (3.20), (3.22) and (3.23). ....	57
3.6	Food Snack Choice Questions .....	59
3.7	Confusion matrix .....	66
3.8	Prediction performance of the 3 models with shuffled labels: the average of the 3 performance measures $A_1, A_2$ and $A_3$ (AUC) taken across the 10 food choice questions for the three competing models $M_1, M_2$ and $M_3$ . For model $M_3$ the choice of $d$ is taken as 8. ....	67
3.9	10-fold cross-validation overall prediction accuracy (in %) for the 10 questions Q1-Q10 corresponding to $d= 4, 5, 6, 7, 8$ for DSSA and ISSA (model $M_3$ ). Significant results (instances of at least a 1% improvement in DSSA) are highlighted in bold. ....	70
3.10	The average of the 3 performance measures $A_1, A_2$ and $A_3$ (AUC) taken across the 10 food choice questions for the three competing models $M_1, M_2$ and $M_3$ . For model $M_3$ the choice of $d$ is taken as 8. ....	74

## 1. INTRODUCTION

Numerous examples of time series data from the real world exhibit a dynamic behavior in terms of its statistical properties leading to nonstationarity. Analyzing time series data that is nonstationary is an important non-trivial problem in statistics that has several interesting applications in areas such as neuroscience, economics, finance and seismology, to name a few. Motivated by these real world applications, different techniques have been proposed to handle problems relating to nonstationarity in time series. In view of modeling temporal dependence and other related tasks such as forecasting, it is often more desirable to work with stationary time series and many of the existing methods aim at reducing the problem of nonstationarity to a stationary time series setting.

The spectral representation of a  $p$ -variate stationary time series  $X_t$  is given by

$$X_t = \int_{-\pi}^{\pi} A(\omega) e^{it\omega} dZ(\omega), \quad (1.1)$$

where  $Z(\omega)$  is a zero-mean  $p$ -dimensional orthogonal increment process on  $[-\pi, \pi]$  and  $A(\omega)$  for  $\omega \in [-\pi, \pi]$  is a  $p \times p$  complex valued matrix such that the  $p \times p$  spectral matrix  $f(\omega)$  of  $X_t$  is given by

$$f(\omega) = A(\omega)A(\omega)^* \quad (1.2)$$

where  $A(\omega)^* = \overline{A(\omega)}^T$  denotes the conjugate transpose. The spectral matrix of the stationary time series contains all the information regarding the covariance structure of the series. An interesting and broad type of nonstationarity is the notion of locally stationary time series developed and studied by Dahlhaus (1997) which allows for sudden changes in the covariance or spectral characteristics of a slowly changing nonstationary process. In this situation the matrix  $A(\omega)$  defined above in (1.1) would be replaced by a time-varying version  $A_t(\omega)$ . Under this assumption

on the observed process, we propose a change point detection method in Chapter 2 that attempts to find time points wherein the statistical properties of the time series changed. Change point detection in nonstationary time series potentially leads to smaller time periods wherein different stationary models can be fit to the data. In Section 1.1 below we review the relevant literature on change point detection and introduce our proposed method for detecting change point locations.

Another related technique for handling nonstationarity is transforming the observed multivariate nonstationary time series to a stationary one via instantaneous linear transformations. The discrete Fourier transform (DFT) of an observed  $p$ -variate time series  $X_t$ ,  $1 \leq t \leq T$ , is given by

$$J_X(\omega_k) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T X_t \exp(-it\omega_k), \quad (1.3)$$

where  $\omega_k = \frac{2\pi}{T}k$ ,  $k = 1, 2, \dots, T$ . Recall that the DFT series is periodic, i.e.  $J_X(\omega_r) = J_X(\omega_{T+r})$ . It is also well known that if  $\{X_t\}$  has a constant mean and is covariance stationary i.e second-order stationary, its DFTs are asymptotically uncorrelated when  $\omega_i \neq \omega_j$ , i.e.

$$\text{cov}(J_X(\omega_i), J_X(\omega_j)) = O\left(\frac{1}{T}\right), \quad (1.4)$$

See Theorem 4.3.1 and Theorem 4.3.2 of Brillinger (2001). It is because of attractive properties like the above that many of the time series methodologies are set in the frequency domain as opposed to the time domain. In Section 1.2 below we introduce stationary subspace analysis (SSA) and our proposed method for finding stationary linear transformations of the observed second-order nonstationary process.

## 1.1 Change Point Detection

Over the past few decades there have been many methods proposed to detect change points in time series and the literature is rather huge, see Killick et al. (2012) for a reference repository. Several applications in various fields such as bioinformatics, signal processing, finance have motivated the use of change point detection methods, see Chen and Gupta (1997), Picard

et al. (2005), Muggeo and Adelfio (2011) and Kirch et al. (2015) for some applications. Despite the widespread applicability, many existing methods impose either independence assumption on the data or allow for dependence and make parametric model assumptions that may lead to misspecification problems.

In the multivariate setting, Aue et al. (2009) detect changes in the covariance matrix across observations using a nonparametric CUSUM type test, and Matteson and James (2014) and Preuss et al. (2015) are two recent methods where the former works with i.i.d data and propose a bisection type procedure to estimate locations of change points. The method in Preuss et al. (2015) is nonparametric and in the spectral domain that uses a moving sum type test statistic involving the infinity norm to quantify differences in spectral matrices in adjacent segments of the observed process. Their technique to identify the locations of change points is similar to Last and Shumway (2008) wherein they sequentially find points of local maximal deviation in the components of their deviance metric. The idea here is to select the point of maximal deviation from an initial set of points and after a selection, the neighbourhood of the selected point is removed from the initial set of points. The procedure is repeated, each time with a reduced set of points, and with suitable stopping criterion the final list of change point locations is determined.

Adak (1998) introduces the notion of piecewise (linear) locally stationary processes which allows for abrupt changes in the covariance characteristics of a slowly varying nonstationary process. We adhere to this type of nonstationarity in the multivariate context and propose a nonparametric method to detecting change point locations. The key features of our method are: (1) Quantifying differences in the spectral matrices between two candidate segments using the *integrated squared Euclidean norm* ( $L_2$  norm) as opposed to the infinity norm used in Preuss et al. (2015). (2) Sequentially identifying points of local maxima in the test statistic. At each identification, we test for the significance of the point being a change point. (3) Relying on the stationary block bootstrap procedure in Politis and Romano (1994) to obtain critical values of the test statistic under the null hypothesis as opposed to the autoregressive bootstrap used in

several other methods including Preuss et al. (2015). The block bootstrap avoids the Gaussian assumption on the observed process, and problems like order selection and fitting high order autoregressive models with a large number of parameters. After detecting change points, we present a new ratio type estimator to identify the components responsible for the change. In addition, a Monte Carlo type construction of confidence intervals for the change points is provided.

In Section 2.8.1, we compare the performance of our method to other recent methods through a simulation study. In Section 2.8.2, we apply our method to identify change points in strong motion seismograph data viewed as a tri-variate time series and also identify the components responsible for significant seismic activity.

## 1.2 Stationary Subspace Analysis (SSA)

The problem of linearly transforming a multivariate nonstationary time series to a stationary time series is of fundamental importance in application areas such as neuroscience and economics. For example, Lemm et al. (2011) and von Büнау et al. (2010) point out that electroencephalogram (EEG) and fMRI signals measuring brain activity appear often as nonstationary time series. Removing the nonstationarity from the observed process is extremely useful for classification purposes in Brain-Computer Interface experiments. In economics one often has to deal with nonstationary data and the powerful idea of *cointegration* analysis, introduced and developed by Granger (1981), Engle and Granger (1987), Johansen (1991), is concerned with finding stationary linear transformations of a unit-root nonstationary time series within the parametric classes of vector-AR (VAR) models. Another related technique in the context of dimension reduction and factor analysis is that of identifying common factors in multivariate time series where the factors could be either stationary or nonstationary. Peña and Poncela (2006) and Nieto et al. (2016) discuss this framework where the observed multivariate process is assumed to be linearly generated by unobserved factors following seasonal VARIMA type models. Other applications include geophysics Hara et al. (2012) and wind energy mitigation Cardinali and Nason (2011).

Stationary subspace analysis (SSA), introduced by von Büнау et al. (2009a), attempts to



find stationary transformations, in lower dimensions, of multivariate nonstationary time series (independent, non-identically distributed random vectors). The model setup for SSA is that the observed  $p$ -variate time series is a linear transformation of (unobserved)  $d$  stationary and  $p - d$  nonstationary sources. The notions of nonstationarity and stationarity used in von Büнау et al. (2009a) and subsequent papers are rather restrictive. The former refers to independent heterogeneous observations and the latter to the time-invariance of the first two moments or the mean and lag-0 covariance of the data; Panknin et al. (2016) allows for stationarity with respect to the first four moments at each time point. The matrix of the linear transformation in SSA is found by dividing the observed time series data into  $N$  segments and minimizing a Kullback-Leibler (KL) divergence between Gaussian distributions measuring differences in the means and covariances across these segments.

In Chapter 3 we describe a frequency domain SSA method for the general class of multivariate second-order nonstationary time series; see Brockwell and Davis (1991), Pourahmadi (2001). The key tools and steps are:

(i) We rely on near uncorrelatedness of the discrete Fourier transform (DFT) of a second-order stationary time series at unequal Fourier frequencies and a test for second-order stationarity, see Dwivedi and Subba Rao (2011) and Jentsch and Subba Rao (2015a), gauging the size of the estimated covariances of the DFTs at various lags. A discrepancy measure is constructed and the demixing matrix is obtained by optimizing this measure.

(ii) In the SSA, von Büнау et al. (2009a), Blythe et al. (2012), a sequential likelihood ratio test is used to determine the dimension of the stationary subspace  $d$  when the observations are independent and normally distributed. We rely on a sequential test of second-order stationarity to determine  $d$  without the independence assumption, and study the consistency of the estimated  $d$  using the asymptotic distribution of the test statistic under the alternative hypothesis of local stationarity of the time series (Dahlhaus (1997), Dahlhaus (2012)).

Section 3.3 includes a simulation study wherein the performance of our method is compared

to that of SSA.

In Section 3.4 an application of our SSA method in neuroeconomic experiments. An emerging literature in neuroeconomics uses brain signals to directly explain choice behavior. One of the models used to explain decision making is the neural random utility model - NRUM; Webb et al. (2013). EEG data have been used to predict purchase decisions Ravaja et al. (2013), consumer's future choices Telpaz et al. (2015), predict preferences Khushaba et al. (2013, 2012) and response to advertisements Boksem and Smidts (2015); Venkatraman et al. (2015).

EEG signals from different electrodes measuring brain activity have, in the past, been regarded as a multi-dimensional nonstationary time series; see Ombao et al. (2005), von Büнау et al. (2010) for examples. Kaplan et al. (2005) regard the nonstationarity as the "unavoidable noise" in the brain signal. Here the nonstationary sources in the brain signal contributes to the noise in the EEG data and removing this nonstationarity is extremely useful for prediction purposes in brain related experiments. We use the words noise and nonstationarity interchangeably because in our setup the nonstationary sources contribute to parts of the signal that are unrelated to the task related activity in the experiment. Hence eliminating nonstationarity reduces noise in the brain signal. See Section 3.4.2 and Figures 3.1, 3.2 for illustrations of the signal before and after noise reduction. von Büнау et al. (2009a) and von Büнау et al. (2010) associate *alpha* oscillations in the data as a nonstationary source. These oscillations appear usually in the range of 8-12 Hz and are associated with blinking, fatigue or tiredness. Over the course of the experiment such changes in the EEG time series are unrelated to the experimental task and corrupt the signal.

We employ the DSSA technique from Section 3.2, as a noise reduction step to separate stationary (useful signal) and nonstationary sources to reduce noise in the EEG brain signal. This is important because using this process may move neurophysiological responses to become more aligned with the design of traditional economics experiments. In other words the nonstationary sources in the brain signal are associated with variations in the mental state that are unrelated to the experimental task at hand. Hence the DSSA technique can be useful in reducing the number

of trials needed from each participant in neuroeconomic experiments. More importantly, the technique greatly improves the prediction performance of an incentivized economic food choice task.

The rest of the report is organized in the following way: Chapters 2, 3 describe in detail the two methodologies, change point detection and stationary subspace analysis, for handling nonstationary time series. Simulation studies are presented and real data applications of the two techniques are discussed. Chapter 4 provides a summary of the proposed methods and lists some of the ongoing and future works.

## 2. CHANGE POINT DETECTION IN PIECEWISE STATIONARY TIME SERIES

In this chapter we describe our nonparametric change point detection method from Sundararajan and Pourahmadi (2018a). We locate points of local maxima in the deviance measure and test for the significance of each of them being change points. Finally, we detect the components responsible for the change at each change point location. The method is developed within the framework of piecewise stationary processes. To begin with, we state the working model which follows a multivariate version of the piecewise stationary processes in Adak (1998).

### 2.1 Working Model and the Approach in Preuss et al. (2015)

Let  $X_{t,T}$ ,  $1 \leq t \leq T$ , be a  $p$ -dimensional zero-mean piecewise stationary time series with piecewise constant spectral density matrix corresponding to break point locations, rescaled to the unit interval,  $0 = b_0 < b_1 < \dots < b_K < b_{K+1} = 1$  where  $K$  is assumed fixed but unknown. Let  $I_k = (b_k, b_{k+1}]$  for  $k = 0, 1, \dots, K$ , then we have

$$X_{t,T} = \int_{-\pi}^{\pi} A_{t,T}(\omega) e^{it\omega} dZ(\omega), \quad (2.1)$$

where (a).  $Z(\omega)$  is a zero-mean  $p$ -dimensional orthogonal increment process on  $[-\pi, \pi]$  and (b). There exists constants  $c_1 \geq 0$ ,  $1/2 < c_2 < 1$  and a  $2\pi$ -periodic matrix valued function  $A : [0, 1] \times [-\pi, \pi] \rightarrow \mathbb{C}$  with  $A(u, \omega) = \overline{A(u, -\omega)}$  and  $A(u, \omega) = \sum_{k=0}^K I(u \in I_k) A_k(\omega)$  such that

$$\max_{t:(t/T) \in I_k} \sup_{\omega} |A_{t,T}(\omega)_{i,j} - A_k(\omega)_{i,j}| \leq c_1 T^{-c_2},$$

for  $i, j = 1, 2, \dots, p$ .  $A_k(\omega)$  is the  $p \times p$  transfer function matrix corresponding to the interval  $I_k$  and  $\sup_{\omega} |A_k(\omega)| < \infty \forall k$ . The time-varying spectral matrix that is piecewise constant w.r.t

time is defined as

$$f\left(\frac{t}{T}, \omega\right) = \sum_{k=0}^K I\left(\frac{t}{T} \in I_k\right) A_k(\omega) A_k(\omega)^* = \sum_{k=0}^K I\left(\frac{t}{T} \in I_k\right) f_k(\omega),$$

where  $A_k(\omega)^* = \overline{A_k(\omega)}^T$  denotes the conjugate transpose and  $f_k(\omega) = A_k(\omega)A_k(\omega)^*$ . The model in (2.1) has an equivalent linear process representation with time-varying coefficients, see Remark 2.2 of Dahlhaus (2000).

For any point  $b \in (0, 1)$ ,  $\delta_b > 0$ ,  $\nu \in [0, 1]$ , Preuss et al. (2015) defined a measure of local change by the  $p \times p$  matrix

$$D_p(b, \nu) = \frac{1}{\delta_b} \left[ \int_{-\nu\pi}^{\nu\pi} \int_b^{b+\delta_b} f(u, \lambda) du d\lambda - \int_{-\nu\pi}^{\nu\pi} \int_{b-\delta_b}^b f(u, \lambda) du d\lambda \right].$$

Their global measure of discrepancy is defined as the supremum of the measure of local change over  $(b, \nu)$ :

$$D_p = \sup_{b, \nu} \|D_p(b, \nu)\|_{\infty}, \quad (2.2)$$

with  $\|\cdot\|_{\infty}$  denoting the infinity norm of the matrix. They utilize a sample version of  $D_p$  based on periodogram matrices as the test statistic for testing the existence of change points. Critical values of the test statistic under the null are obtained by a vector-autoregressive (VAR) bootstrap where the order increases with increasing sample size. Once there is a rejection for this test, the locations of change points are determined by setting thresholds for the entries of the matrix  $\sup_{\nu \in [0, 1]} |D_p(b, \nu)|$ . This is done by first identifying an initial set of candidate points and picking a subset of points for which the entries of the discrepancy matrix exceed a prescribed threshold,

$$N^{\gamma} \sup_{\nu \in [0, 1]} |(D_p(b, \nu))_{(i,j)}| > \epsilon_{(i,j)}(b), \quad (2.3)$$

for some point  $b$ , where  $N$  is the size of a local neighbourhood around  $b$ ,  $\gamma$  is a tuning parameter and  $\epsilon_{(i,j)}(b)$  is the threshold.

## 2.2 An Alternative Measure of Local Change

We propose a measure of local change around a point, say  $b$ , by quantifying the vectorized (vec) form of the difference of the spectral density matrices of the data on either side of the point using *integrated squared Euclidean norm* ( $L_2$  norm). More precisely, we use the metric

$$D(b) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\text{vec}[f_L(b, \omega) - f_R(b, \omega)]\|^2 d\omega, \quad (2.4)$$

where  $\|\cdot\|$  is the Euclidean norm and  $f_L(b, \omega)$  and  $f_R(b, \omega)$  are spectral matrices over the left and right neighborhoods of  $b$ , i.e.  $(\lfloor bT \rfloor - N, \lfloor bT \rfloor]$  and  $(\lfloor bT \rfloor, \lfloor bT \rfloor + N]$ , respectively, and  $N > 0$  determines the length of a local neighbourhood.

In our preliminary data analysis we had noticed that the squared Euclidean norm in (2.4) was able to distinguish actual change point locations in a more decisive way than the statistic in (2.2) for small sample sizes. In addition, spectral density based test statistics involving the  $L_2$  norm have been shown to be consistent; see Paparoditis (2000), Eichler (2008), Dette and Paparoditis (2009), Jentsch and Subba Rao (2015a) where hypothesis tests about spectral densities are carried out. Simulation techniques are implemented to obtain critical values of the tests to overcome the slow rate of convergence of such test statistics to a normal distribution. See Sundararajan and Pourahmadi (2018a) for additional details.

### 2.2.1 Estimating the Measure of Local Change

In this section we present an estimated form of the metric in (2.4) that will be used in our method for an arbitrary point  $b \in (0, 1)$ .

Recall that the discrete Fourier transform (DFT) and the periodogram of a  $p$ -variate series  $X_1, X_2, \dots, X_N$  are defined as

$$J(\omega) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N X_t \exp(-it\omega), \quad I_N(\omega) = J(\omega)J(\omega)^*,$$

where  $J(\omega)^*$  denotes the conjugate transpose, and the estimated  $p \times p$  spectral density matrix, for  $\omega \in [-\pi, \pi]$  is given by

$$\hat{f}(\omega) = \frac{1}{N} \sum_{j=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} K_h(\omega - \omega_j) I_N(\omega_j),$$

where  $\omega_j = \frac{2\pi}{N}j$  and  $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$  where  $K(\cdot)$  is a nonnegative symmetric kernel function and  $h$  denotes the bandwidth. The regularity conditions on the kernel and bandwidth are enforced as in Eichler (2008) to ensure uniform consistency in  $\omega \in [-\pi, \pi]$  of the estimated spectral matrices, and is further discussed in Section 2.7. Thus, an estimate of the metric in (2.4) is given by

$$\hat{D}(b) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\text{vec}[\hat{f}_L(b, \omega) - \hat{f}_R(b, \omega)]\|^2 d\omega, \quad (2.5)$$

where  $\hat{f}_L(b, \omega_k)$ ,  $\hat{f}_R(b, \omega_k)$  are the estimated spectral density matrices using the data over the two local neighborhoods  $(\lfloor bT \rfloor - N, \lfloor bT \rfloor]$  and  $(\lfloor bT \rfloor, \lfloor bT \rfloor + N]$ , respectively.

We choose  $N = 2^{g(T)}$  as a dyadic positive integer where  $g(T)$  is an integer sequence monotone in the sample size  $T$ . This choice of  $N$  is helpful in faster computations through the fast Fourier transform. A discretized version of (2.5) is

$$\hat{D}(b) = \frac{1}{N} \sum_{k=1}^N \|\text{vec}[\hat{f}_L(b, \omega_k) - \hat{f}_R(b, \omega_k)]\|^2, \quad (2.6)$$

where  $\omega_k = (2\pi/N)k$ ,  $k = 1, 2, \dots, N$ , are the fundamental Fourier frequencies.

### 2.3 Finding Change Point Locations

To locate significant change points we first compute  $\hat{D}(u)$  for every  $u \in \{N/T, (N+1)/T, \dots, (T-N)/T\}$  where  $N$  determines the size of a neighbourhood around potential change points. Then the point, say  $u^*$ , giving the maximum value of  $\hat{D}(\cdot)$  over the set  $\{N/T, (N+1)/T, \dots, (T-N)/T\}$  is located and tested for significance using the test described in Section 2.4. If found significant  $u^*$  is included as a significant change point and  $\hat{D}(u)$  for every

$u \in (u^* - N/T, u^* + N/T)$  is set to zero. This procedure is repeated until the identified point  $u^*$  is found insignificant or when  $\hat{D}(u^*) = 0$ . This technique to estimate the location of change points is given below in Algorithm 1.

---

**Algorithm 1:** Detecting locations of change points

---

**Result:** Output  $P$ : the set of significant change points

Initialize  $P = \text{NULL}$ ,  $\text{Flag}=0$ ;

**while**  $\text{Flag}=0$  **do**

Find  $u^*$  where  $u^* = \arg. \max_{u \in \mathcal{C}} \hat{D}(u)$  where  $\mathcal{C} = \{N/T, (N+1)/T, \dots, (T-N)/T\}$ ;

Test whether  $u^*$  is significant change point;

**if**  $\hat{D}(u^*) = 0$  or  $u^*$  is NOT significant **then**

|  $\text{Flag}=1$ ;

**else**

| Set  $\hat{D}(u) = 0$  for  $u \in (u^* - N/T, u^* + N/T)$ ;

|  $P = P \cup u^*$ ;

**end**

**end**

---

### 2.3.1 Choice of $N$ : Length of Local Neighborhood

The choice of  $N$  depends on the nature of changes in the covariance. Smaller changes require large values of  $N$  to get identified while larger changes can be detected with smaller values of  $N$ . As pointed out by a referee, we adopt a multiscale procedure to select the appropriate value of  $N$ . Following the technique given in Section 4.2 of Messer et al. (2014), we consider a set of  $n$  choices for  $N$  given by  $\sqrt{T} < N_1 < N_2, \dots < N_n < T^{5/6}$ . In addition, each choice  $N_i$ ,  $i = 1, 2, \dots, n$ , is dyadic and we have

$$N_i = 2^{l_i} \text{ where } l_i = \lfloor \log_2(\sqrt{T}) \rfloor + i \quad (2.7)$$



for  $i = 1, 2, \dots, n - 1$  and  $l_n = \lfloor \log_2(T^{5/6}) \rfloor$ . Let  $P_i$  denote the set of change points estimated using Algorithm 1 with neighborhood length choice  $N_i$ . Set  $P = P_1$  where  $P$  denotes the final set of estimated change points by our method. For any point  $v \in P_2$ ,  $v$  is added to the set  $P$  only if it does not belong to a  $N_2$ -neighborhood of any of the existing points in the set  $P$ . The procedure is successively moved forward till neighborhood choice  $N_n$ .

The above technique is utilized while determining the change point locations for the various models simulated in Section 2.8.1.1 and also in the application discussed in Section 2.8.2.

For a single neighborhood length choice  $N$ , we take the largest value of  $N \in \{N_1, N_2, \dots, N_n\}$  for which there is an addition of a point to the set  $P$  in the procedure described above. More precisely  $N = N_{i^*}$  where  $i^*$  is the largest value in the set  $\{1, 2, \dots, n\}$  for which the iterative procedure described above adds a point to the set  $P$  during iteration  $i^*$  (i.e with neighborhood length choice  $N_{i^*}$ ). In case there is no point added to the set  $P$  for any choice  $N_i$ ,  $i = 1, 2, \dots, n$ , we set  $N = N_n$ . This technique leading to a single neighborhood length is utilized while constructing confidence intervals for change points in Sections 2.6, 2.8.1.4 and also in Sections 2.8.1.2, 2.8.1.3 that involve estimation of the empirical size and power of the test for various simulation examples.

## 2.4 Testing for Significant Change Points

Here we describe a test procedure for deciding the significance of a change point.

Given any  $b \in (0, 1)$ , a point of local maxima, we consider the statistic in (2.6) which serves as the test statistic. Under the null hypothesis  $H_0$  that  $b$  is away from the neighbourhood of any of the true change points, the theoretical properties of the test statistic are spelled out in Section 2.7. The limiting distribution of  $\hat{D}(b)$  under the null is derived assuming that  $N/T \rightarrow c$  for some  $0 < c < 1$  as  $T \rightarrow \infty$ . It will also be shown that under the alternative hypothesis, the test based on  $\hat{D}(b)$  is consistent i.e  $P(\hat{D}(b) > A) \rightarrow 1$  for some  $A > 0$ .

We use the block bootstrap technique for stationary processes in Politis and Romano (1994) to obtain the p-value of the test and this is discussed in the next subsection. Here, we avoid the use of autoregressive bootstrap and the associated problem of order selection which may involve

estimation of a large number of parameters.

### 2.4.1 Block Bootstrap for Stationary Processes

Methods for generating realizations of a stationary process include the autoregressive bootstrap Kreiss (1992); Meyer and Kreiss (2015), spectral density based techniques Sun and Chaika (1997); Hu and Schiehlen (1997); Azimmohseni et al. (2015) and the circulant embedding method Percival and Constantine (2006); Helgason et al. (2011), to name a few. We rely on the block bootstrap technique of Politis and Romano (1994) for generating observations from a stationary process. This procedure does not assume Gaussianity in the observed process and involves the choice of one tuning parameter namely the block length.

Given  $T_0$  observations  $X_1, X_2, \dots, X_{T_0}$ , let  $B_{i,L} = \{X_i, X_{i+1}, \dots, X_{i+L-1}\}$  denote a block of length  $L$ . Note that when  $j > T_0$ ,  $X_j = X_{j(\text{mod } T_0)}$ . Let  $L_i, i = 1, 2, \dots$ , be i.i.d random variables from a geometric distribution with mean  $1/q$  i.e  $P(L_i = a) = q(1 - q)^{a-1}$  for  $a = 1, 2, \dots$ , and let  $U_k, k = 1, 2, \dots$ , be i.i.d random variables from a discrete uniform distribution over the set  $\{1, 2, \dots, T_0\}$ . Now the  $r^{\text{th}}$  bootstrap sample,  $r = 1, 2, \dots, B$ , is given by  $\{X_1^{r,*}, X_2^{r,*}, \dots, X_{T_0}^{r,*}\}$  where the first  $L_1$  observations are given by block  $B_{U_1, L_1}$ , the next  $L_2$  observations are given by block  $B_{U_2, L_2}$  and so on. The procedure is terminated once  $T_0$  observations are obtained in the bootstrap sample. After obtaining  $B$  such bootstrap samples, we compute the p-value of the test as  $\frac{1}{B} \#\{\hat{D}_r^*(b) > \hat{D}_{\text{observed}}(b), 1 \leq r \leq B\}$  where  $\hat{D}_r^*(b)$  and  $\hat{D}_{\text{observed}}(b)$  are estimates of  $\hat{D}(b)$  based on the  $r^{\text{th}}$  bootstrap sample and the original sample respectively.

The choice of the random block lengths is determined by the probability  $q$  of the geometric distribution. Here we resort to the procedure in Politis and White (2004); Patton et al. (2009) where under certain mixing conditions on the observed process, the empirical block length selection is shown to be optimal. However, this procedure is for the univariate case. Hence we apply it to each component of the multivariate process and obtain the block length as the average over all components as in Jentsch and Subba Rao (2015b).

## 2.5 Identifying the Components Responsible for the Change

Once the break point locations  $\{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_{\hat{K}}\}$  are estimated, the components responsible for the change in the covariance structure are identified as follows. For any  $\hat{b} \in \{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_{\hat{K}}\}$ , we consider the  $p \times p$  matrix

$$M_{\hat{b}} = \frac{1}{N} \sum_{k=1}^N [\hat{f}_L(\hat{b}, \omega_k) - \hat{f}_R(\hat{b}, \omega_k)]. \quad (2.8)$$

where  $\omega_k = \frac{2\pi}{N}k$ . For  $i \leq j$ , we arrange the absolute value of the entries given by  $|M_{\hat{b},ij}|$  in descending order, say  $E_1 \geq E_2 \geq \dots \geq E_{p_0}$ , where  $p_0 = p(p+1)/2$ . Finally we conclude that the components corresponding to the first  $r$  entries have a significant contribution to the change where

$$r = \arg. \max_{1 \leq j \leq \zeta \times p_0} \frac{E_j}{E_{j+1}} \quad (2.9)$$

where  $0 < \zeta < 1$  is chosen to avoid small values of  $E_j$  resulting in large values of the ratio. An illustration of this can be seen in our application presented in Section 2.8.2 with  $\zeta = 0.6$ .

## 2.6 Constructing Confidence Intervals for Change Points

Constructing confidence intervals for change points has been discussed recently in Yau and Zhao (2015) for univariate time series. Here, simultaneous confidence intervals for multiple change points under a piecewise autoregressive process assumption on the observed data are constructed. They utilize a likelihood ratio based statistic defined over local neighborhoods for measuring the discrepancy at candidate points and implement a minimum description length based model selection approach to estimate change points. Here we propose a Monte Carlo type construction of confidence intervals for true change points.

Let  $P = \{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_{\hat{K}}\}$  be the set of estimated change point locations and let  $\hat{b}_0 = 1$  and  $\hat{b}_{\hat{K}+1} = T$ . For any point  $\hat{b}_j \in P$ , the consistency results along with the assumptions given in Section 2.7 imply that there is only one change point in  $(\hat{b}_j - N/T, \hat{b}_j + N/T]$ . We now define

$\bar{b}_j$  as

$$\bar{b}_j = \arg. \max_{v \in I_{\hat{b}_j}} \hat{D}\left(\frac{v}{T}\right), \quad (2.10)$$

where  $I_{\hat{b}_j} = I_{\hat{b}_j,L} \cup I_{\hat{b}_j,R}$ . Here  $I_{\hat{b}_j,L}$  and  $I_{\hat{b}_j,R}$  denote two intervals corresponding to  $[\hat{b}_j T] - N \leq t \leq [\hat{b}_j T]$  and  $[\hat{b}_j T] + 1 \leq t \leq [\hat{b}_j T] + N$  respectively. We now construct  $B$  intervals as follows

$$X_{t,\hat{b}_j}^* = I\left(t \in I_{\hat{b}_j,L}\right) U_{t,\hat{b}_j}^* + I\left(t \in I_{\hat{b}_j,R}\right) V_{t,\hat{b}_j}^*$$

for  $[\hat{b}_j T] - N \leq t \leq [\hat{b}_j T] + N$ . Here  $U_{t,\hat{b}_j}^*$ ,  $t \in I_{\hat{b}_j,L}$ , constitutes a block bootstrap sample generated from observations in  $I_{\hat{b}_j,L}$  using the technique described earlier in Section 2.4.1. Similarly we have  $V_{t,\hat{b}_j}^*$ ,  $t \in I_{\hat{b}_j,R}$ , constituting a block bootstrap sample generated from observations in  $I_{\hat{b}_j,R}$ .

We thus obtain  $\bar{b}_j^{(s)}$ ,  $s = 1, 2, \dots, B$ , where  $\bar{b}_j^{(s)}$  denotes  $\bar{b}_j$  for the  $s^{th}$  simulated interval. Using this we obtain the  $100(1 - \alpha)\%$  confidence interval for the true change point  $b_j$  as  $[\bar{b}_{j,\alpha/2}, \bar{b}_{j,1-\alpha/2}]$  where  $\bar{b}_{j,\alpha/2}$  denotes the  $(\alpha/2)^{th}$  sample quantile of  $\bar{b}_j^{(1)}, \bar{b}_j^{(2)}, \dots, \bar{b}_j^{(B)}$ .

Regarding the choice of  $N$  used for constructing these intervals, we refer back to the procedure described in Section 2.3.1 and select the largest value of  $N \in \{N_1, N_2, \dots, N_n\}$  for which there is an addition of a point to the set  $P$  in the iterative procedure.

## 2.7 Theoretical Properties

In this section we establish the asymptotic properties of the test statistic  $\hat{D}(b)$  for  $b \in (0, 1)$  and show that the corresponding test is consistent. The key assumptions are that  $N = O(T)$  and for some  $0 < c < 1$ , consecutive change points are at least  $c$  units apart on the rescaled time interval of  $[0, 1]$ .

Let  $S_{1,b} = ([bT] - N, [bT])$ ,  $S_{2,b} = ([bT], [bT] + N)$  and  $Y_t^{(b)} = (Z_{1,t}^{(b)}, Z_{2,t}^{(b)})^T$  where for  $i = 1, 2$ ,  $\{Z_{i,1}^{(b)}, Z_{i,2}^{(b)}, \dots, Z_{i,N}^{(b)}\} = \{X_t : t \in S_{i,b}\}$ . Now for  $Y_t^{(b)} \in \mathbb{R}^{2p}$ , the  $2p \times 2p$  partitioned spectral matrix is

$$f^{(b)}(\omega) = \begin{bmatrix} f_{11}^{(b)}(\omega) & f_{12}^{(b)}(\omega) \\ f_{21}^{(b)}(\omega) & f_{22}^{(b)}(\omega) \end{bmatrix},$$

where  $f_{11}^{(b)}(\omega) = f_L(b, \omega)$  and  $f_{22}^{(b)}(\omega) = f_R(b, \omega)$ ,  $f_L(b, \omega)$  and  $f_R(b, \omega)$  defined earlier, are the  $p \times p$  spectral matrices of  $X_t$  over the intervals  $S_{1,b}$  and  $S_{2,b}$ , respectively, and  $f_{12}^{(b)}(\omega)$  is the cross-spectrum of  $Z_{1,t}^{(b)}$  and  $Z_{2,t}^{(b)}$ . It must be noted that joint stationarity of  $(Z_{1,t}^{(b)}, Z_{2,t}^{(b)})^T$  will hold due to linearity of the observed process  $\{X_t\}$ . The choices of the kernel function and bandwidth for obtaining a consistent estimator  $\hat{f}^{(b)}(\omega)$  of  $f^{(b)}(\omega)$  is the same as in Section 2.2 satisfying Assumption 2 below.

We now introduce the assumptions necessary for obtaining the results.

**Assumption 1.** Let  $\{Y_t^{(b)}\}$ ,  $t \in \mathbb{Z}$  be a  $2p$ -variate zero-mean stationary process. For any  $k > 0$ , the  $k^{\text{th}}$  order cumulants of  $Y_t^{(b)}$  satisfy

$$\sum_{u_1, u_2, \dots, u_{k-1} \in \mathbb{Z}} [1 + |u_j|^2] |c_{a_1, a_2, \dots, a_k}(u_1, u_2, \dots, u_{k-1})| < \infty$$

for  $j = 1, 2, \dots, k-1$  and  $a_1, a_2, \dots, a_k = 1, 2, \dots, 2p$  where  $c_{a_1, a_2, \dots, a_k}(u_1, u_2, \dots, u_{k-1})$  is the  $k^{\text{th}}$  order joint cumulant of  $Y_{a_1, u_1}^{(b)}, \dots, Y_{a_{k-1}, u_{k-1}}^{(b)}, Y_{a_k, 0}^{(b)}$  as defined in Brillinger (2001).

This is the same as Assumption 3.1 of Eichler (2008) and requires existence of all order moments of  $Y_t^{(b)}$ .

**Assumption 2.** (a). The kernel function  $K(\cdot)$  is bounded, symmetric, nonnegative and Lipschitz-continuous with compact support  $[-\pi, \pi]$  and

$$\int_{-\pi}^{\pi} K(\omega) d\omega = 1.$$

where  $K(\omega)$  has a continuous Fourier transform  $k(u)$  such that

$$\int k^2(u) du < \infty \text{ and } \int k^4(u) du < \infty.$$

(b). The bandwidth  $h$  is such that  $h^{9/2}N \rightarrow 0$  and  $h^2N \rightarrow \infty$  as  $N \rightarrow \infty$ .

This assumption is needed to ensure uniform consistency of the estimated spectral matrices and ensure sufficient reduction in the bias to not affect the asymptotic convergence of the test statistic; see Assumption 3.3 of Eichler (2008) for details.

Now, let the true change point locations be given by  $0 < b_1 < b_2, \dots < b_K < 1$ . We define  $\mathcal{P}$  as the set of all points in the immediate neighbourhoods of the  $K$  change points,

$$\mathcal{P} = \bigcup_{i=1}^K \{b_i - N/T, \dots, b_i + N/T\}. \quad (2.11)$$

We will now establish two results that describe the behaviour of  $\hat{D}(b)$  when  $b \in \mathcal{P}$  or  $\overline{\mathcal{P}}$  where  $\overline{\mathcal{P}} = \{N/T, (N+1)/T, \dots, (T-N)/T\} \setminus \mathcal{P}$ .

**Theorem 2.7.1.** (Asymptotic Normality) Suppose that Assumptions 1,2 are satisfied. Then for  $b \in \overline{\mathcal{P}}$  with  $N/T \rightarrow c$  for some  $0 < c < 1$  as  $T \rightarrow \infty$  we have

$$2\pi N h^{1/2} \hat{D}(b) - \frac{\mu_0}{\sqrt{h}} \xrightarrow{D} N(0, \sigma_0^2) \quad (2.12)$$

where

$$\mu_0 = A_K \int_{-\pi}^{\pi} \left( \sum_{j_1, j_2=1}^2 (-1 + 2\delta_{j_1 j_2}) |tr(f_{j_1 j_2}^{(b)}(\omega))|^2 \right) d\omega \quad (2.13)$$

and

$$\sigma_0^2 = B_K \int_{-\pi}^{\pi} \left( \sum_{j_1, j_2, j_3, j_4=1}^2 (-1 + 2\delta_{j_1 j_2}) (-1 + 2\delta_{j_3 j_4}) |tr(f_{j_1 j_3}^{(b)}(\omega) \overline{f_{j_2 j_4}^{(b)}(\omega)})|^2 \right) d\omega. \quad (2.14)$$

where  $\xrightarrow{D}$  denotes convergence in distribution,  $A_K = \int_{-\pi}^{\pi} K^2(\omega) d\omega$ ,  $B_K = 4 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} K(\omega) K(\omega + \eta) d\omega d\eta$ ,  $\delta_{rs} = I(r = s)$  is the Kronecker delta and  $tr(\cdot)$  denotes the trace of a matrix.

*Proof.* See Appendix A for details of the proof. □

The entities  $\mu_0$  and  $\sigma_0^2$  are unknown and consistent estimators of both can be obtained by substituting  $f_{ij}^{(b)}$ ,  $i, j \in \{1, 2\}$  with  $\hat{f}_{ij}^{(b)}$ ; see Remark 3.7 of Eichler (2008). Note that  $\hat{f}_{ij}^{(b)}$  can be

treated as mean square consistent estimators with rate  $\sqrt{Nh}$ ; see Equation 5 of Eichler (2008). Thus we have  $\hat{\mu}$  and  $\hat{\sigma}^2$  where  $\hat{\mu} \xrightarrow{P} \mu_0$  and  $\hat{\sigma}^2 \xrightarrow{P} \sigma_0^2$ . Then by Slutsky's theorem we have with  $N/T \rightarrow c$  as  $T \rightarrow \infty$ ,

$$Q_N = \frac{2\pi N\hat{D}(b) - \hat{\mu}/h}{\hat{\sigma}/\sqrt{h}} \xrightarrow{D} N(0, 1) \quad (2.15)$$

when  $b \in \overline{\mathcal{P}}$ . Consistent estimators  $\hat{\mu}$  and  $\hat{\sigma}^2$ , under the data distribution, are required to study the properties of the test statistic when  $b \in \mathcal{P}$ . Now we establish consistency of the test as a consequence of the following theorem.

**Theorem 2.7.2. (Consistency)** *Suppose that Assumptions 1,2 are satisfied. Then for  $b \in \mathcal{P}$  with  $N/T \rightarrow c$  for some  $0 < c < 1$  as  $T \rightarrow \infty$  we have*

$$\frac{Q_N}{N\sqrt{h}} \xrightarrow{P} A_1 \quad \text{and} \quad 2\pi\hat{D}(b) \xrightarrow{P} A_2 \quad (2.16)$$

where  $\xrightarrow{P}$  denotes convergence in probability,  $A_1$  and  $A_2$  are positive constants.

*Proof.* See Appendix A for details of the proof. □

A direct consequence of  $2\pi\hat{D}(b) \xrightarrow{P} A_2$  for some  $A_2 > 0$  is that when  $b \in \mathcal{P}$ ,  $\lim_{N \rightarrow \infty} P(\hat{D}(b) > A) = 1$  for some  $A > 0$ . This implies rejecting the null hypothesis for large values of the test statistic  $\hat{D}(b)$  will result in a consistent test. Also, if there are multiple change point locations  $\{b_1, b_2, \dots, b_K\}$ , Theorem 3.2 implies that  $P(\bigcap_{i=1}^K \{\hat{D}(b_i) > A_1\})$ , for some  $A_1 > 0$ , converges to 1 as  $N \rightarrow \infty$ .

It follows from Theorems 2.7.1, 2.7.2 mean that  $\hat{D}(b) = o_p(1)$  if  $b \in \overline{\mathcal{P}}$ , and  $\hat{D}(b) > A$  for some  $A > 0$  if  $b \in \mathcal{P}$ . Recall that the true measure of local change  $D(\cdot)$  attains local maxima at the true change points  $\{b_1, b_2, \dots, b_K\}$  and equals 0 for points away from the neighbourhoods of change points. Thus we have

$$\sup_{b \in (0,1)} |\hat{D}(b) - D(b)| = o_p(1) \quad (2.17)$$

Therefore, by similar arguments from Theorem 3.6 and Property 3.9 of Preuss et al. (2015), the property given above in (2.17) will imply that  $\hat{b}_i \xrightarrow{P} b_i$  for  $i = 1, 2, \dots, K$  and we will also have  $\hat{K} \rightarrow K$ .

## 2.8 Illustrative Examples

### 2.8.1 Simulation Study

In this section we illustrate the performance of our method in comparison to other recent methods through a few simulation examples, analyze the power of the proposed test and also illustrate the construction of confidence limits for change points. The length of the series is taken as  $T = 256, 512, 1024$ . The probability parameter  $q$  for generating the random block lengths in the bootstrap procedure from Section 2.4.1 is taken according to Politis and White (2004); Patton et al. (2009). The number of bootstrap replications  $B$  is set to 500. For the estimated spectral matrix, our method was implemented using the Bartlett-Priestley kernel with bandwidth  $h = N^{-0.4}$  and the Daniell kernel, see Example 10.4.1 in Brockwell and Davis (1991) with  $m = \sqrt{N}$ . Similar results were obtained for the two kernel choices and only the results from the latter are presented. The significance level for testing for significant change points is set at 1%. A correction could be applied to the choice of significance level as the method involves multiple tests with the number of tests depending on the unknown number of change points. Controlling the overall error rate in this multiple testing scenario requires further theoretical investigation.

First we investigate the ability of our method in detecting actual change point locations as compared to recent methods of Preuss et al. (2015) and Matteson and James (2014). The latter work includes a change point detection method for multivariate piecewise i.i.d data. Here a bisection type procedure that uses a characteristic function based divergence measure is proposed.

#### 2.8.1.1 Example 1: Change Point Detection Accuracy

We conducted 500 simulation runs of each model given below and a histogram of detected change points based on our method is obtained and presented on a 0-100 scale. Likewise, a



similar histogram is constructed for each of the competing methods. In our method, for the length of the dyadic neighbourhood  $N$ , we use the technique in Section 2.3.1.

(a). A bivariate piecewise i.i.d model with one change point at  $b_1 = \{0.5\}$

$$X_{t,T} = I(t/T \leq 0.5)u_t + I(t/T > 0.5) v_t \quad (2.18)$$

where  $u_t$  and  $v_t$  are i.i.d  $N(0, I_2)$  and  $N(0, 3I_2)$  respectively.

(b). A bivariate piecewise i.i.d model with two change points at  $b_1 = 0.3, b_2 = 0.75$

$$X_{t,T} = I(t/T \leq 0.3)u_t + I(0.3 < t/T < 0.75) v_t + I(t/T > 0.75) w_t \quad (2.19)$$

where  $u_t, v_t$  and  $w_t$  are i.i.d  $N(0, I_2), N(0, 2I_2)$  and  $N(0, 3I_2)$ , respectively.

(c). A bivariate AR(1) model with two change points at  $b_1 = 0.5, b_2 = 0.75$

$$\begin{aligned} X_{t,T} = & I\left(\frac{t}{T} \leq 0.5\right) \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} X_{t-1,T} + I\left(0.5 < \frac{t}{T} \leq 0.75\right) \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix} X_{t-1,T} \\ & + I\left(\frac{t}{T} > 0.75\right) \begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix} X_{t-1,T} + \epsilon_t \end{aligned} \quad (2.20)$$

where  $\epsilon_t$  are i.i.d  $N(0, I_2)$ .

(d). A bivariate MA(1) model with two change points at  $b_1 = 0.5$ ,  $b_2 = 0.75$ ,

$$\begin{aligned}
X_{t,T} = & I\left(\frac{t}{T} \leq 0.5\right) \begin{bmatrix} 1.4 & 0 \\ 0 & 0.9 \end{bmatrix} Z_{t-1} + I\left(0.5 < \frac{t}{T} \leq 0.75\right) \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} Z_{t-1} \\
& + I\left(\frac{t}{T} > 0.75\right) \begin{bmatrix} 2.2 & 0.1 \\ -0.3 & 1.3 \end{bmatrix} Z_{t-1} + Z_t
\end{aligned} \tag{2.21}$$

where  $Z_t$  are i.i.d  $N(0, I_2)$ .

(e). A bivariate AR(2) model with two change points at  $b_1 = 0.5$ ,  $b_2 = 0.75$ ,

$$\begin{aligned}
X_{t,T} = & I\left(\frac{t}{T} \leq 0.5\right) \left\{ \begin{bmatrix} 1.1 & 0 \\ 0 & 1.1 \end{bmatrix} X_{t-1,T} + \begin{bmatrix} -0.7 & 0 \\ 0 & -0.7 \end{bmatrix} X_{t-2,T} \right\} \\
& + I\left(0.5 < \frac{t}{T} \leq 0.75\right) \left\{ \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix} X_{t-1,T} + \begin{bmatrix} -0.7 & 0 \\ 0 & 0 \end{bmatrix} X_{t-2,T} \right\} \\
& + I\left(\frac{t}{T} > 0.75\right) \left\{ \begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix} X_{t-1,T} + \begin{bmatrix} 0 & 0 \\ 0 & -0.7 \end{bmatrix} X_{t-2,T} \right\} + \epsilon_t
\end{aligned} \tag{2.22}$$

where  $\epsilon_t$  are i.i.d zero-mean Gaussian with  $\text{cov}(\epsilon_{t_1}, \epsilon_{t_2}) = 0.5$ .

(f). A tri-variate MA(1) model with three change points at  $b_1 = 0.25$ ,  $b_2 = 0.5$ ,  $b_3 = 0.75$ ,

$$\begin{aligned}
X_{t,T} = & I\left(\frac{t}{T} \leq 0.25\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} Z_{t-1} + I\left(0.25 < \frac{t}{T} \leq 0.5\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} Z_{t-1} \\
& + I\left(0.5 < \frac{t}{T} \leq 0.75\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} Z_{t-1} + I\left(\frac{t}{T} > 0.75\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} Z_{t-1} + Z_t
\end{aligned} \tag{2.23}$$

where  $Z_t$  are i.i.d  $N(0, I_3)$ .

(g). Finally, we consider a bivariate time-varying coefficients AR(1) model similar to Preuss et al. (2015) with one change point at  $b_1 = 0.5$  wherein the model assumptions in (2.1) and the assumptions for the theoretical results in Section 2.4 are violated:

$$\begin{aligned}
X_{t,T} = & I\left(\frac{t}{T} \leq 0.5\right) \begin{bmatrix} 0.8 - 3.2(t/T) & 0.3 \\ -0.6 & 0.1 \end{bmatrix} X_{t-1,T} + I\left(\frac{t}{T} > 0.5\right) \begin{bmatrix} 2.4 - 3.2(t/T) & 0.1 \\ -0.6 & 0.1 \end{bmatrix} \times \\
& X_{t-1,T} + \epsilon_t
\end{aligned} \tag{2.24}$$

where  $\epsilon_t$  are i.i.d zero-mean Gaussian with  $\text{cov}(\epsilon_{t,1}, \epsilon_{t,2}) = 0.5$ .

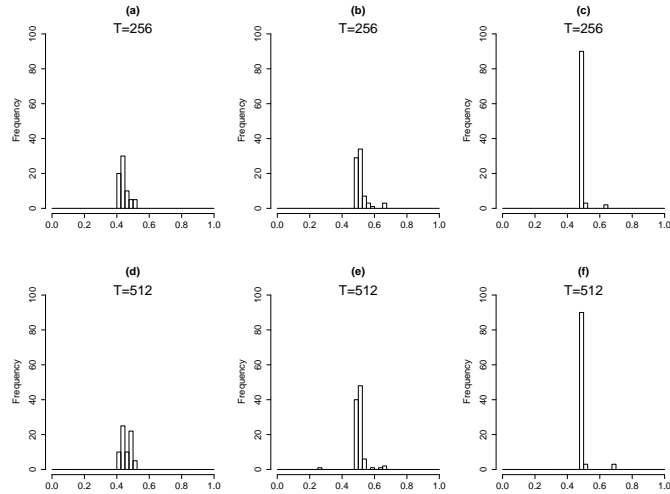


Figure 2.1: Change point at  $\{0.5\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.18) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014).

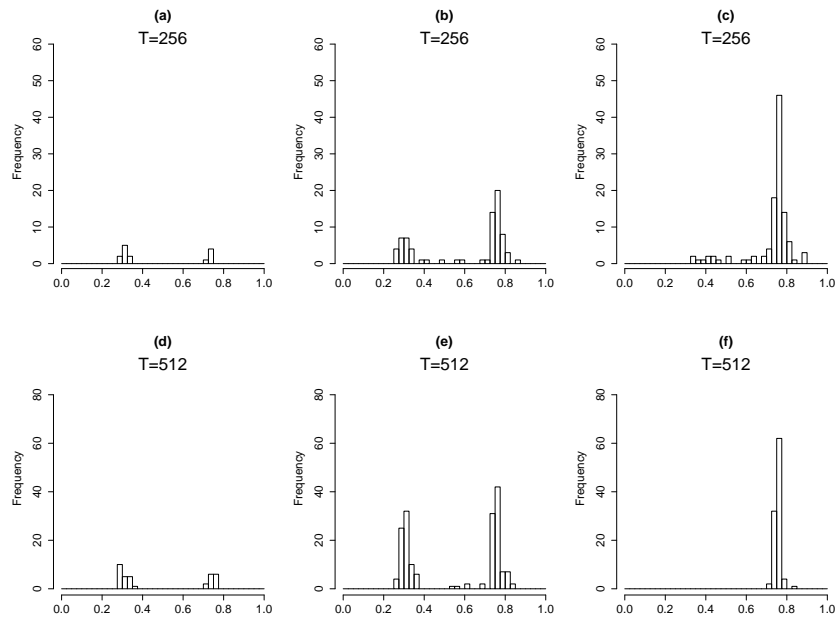


Figure 2.2: Change points at  $\{0.3, 0.75\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.19) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014).

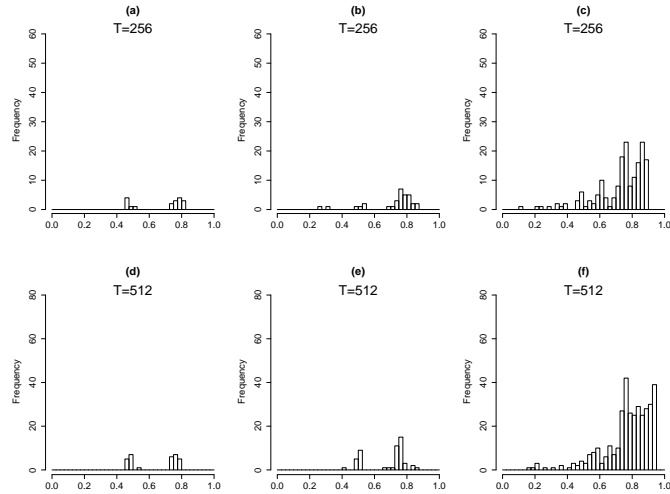


Figure 2.3: Change points at  $\{0.5, 0.75\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.20) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014).

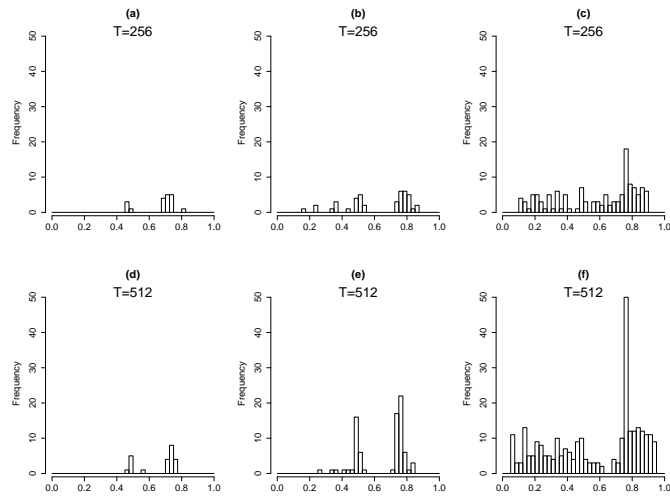


Figure 2.4: Change points at  $\{0.5, 0.75\}$ . (a),(d) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.21) using Preuss et al. (2015) and (b),(e) is the same histogram using our method at 1% level and (c),(f) is the same histogram using Matteson and James (2014).

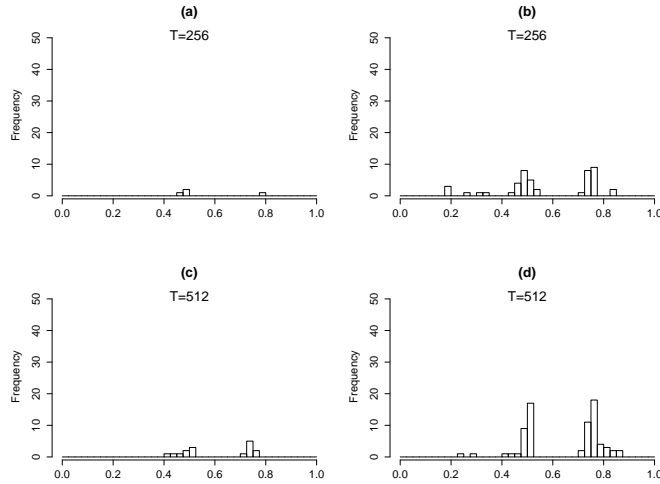


Figure 2.5: Change points at  $\{0.5, 0.75\}$ . (a),(c) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (3.20) using Preuss et al. (2015) and (b),(d) is the same histogram using our method at 1% level.

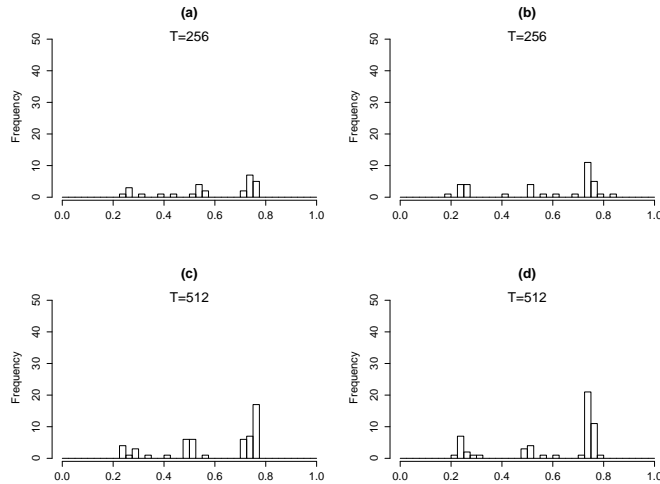


Figure 2.6: Change points at  $\{0.25, 0.5, 0.75\}$ . (a),(c) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.23) using Preuss et al. (2015) and (b),(d) is the same histogram using our method at 1% level.

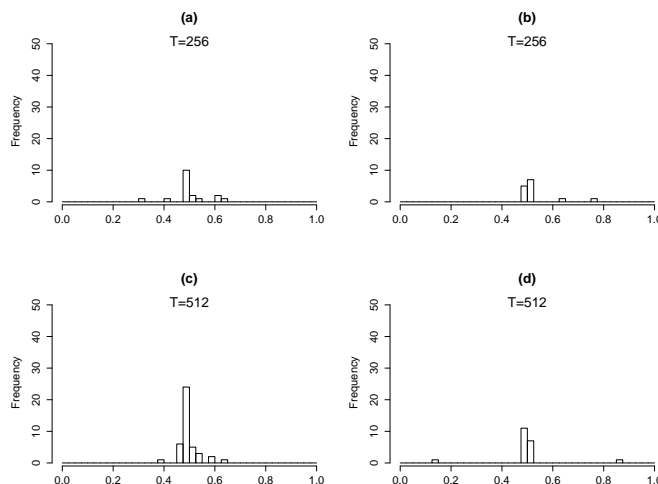


Figure 2.7: (a),(c) is the histogram of detected change point locations (presented on a 0-100 scale) of the model in (2.24) using Preuss et al. (2015) and (b),(d) is the same histogram using our method at 1% level.

In the panel histograms provided in Figures 2.1–2.4, plots (a),(c) is the histogram of detected change point locations using Preuss et al. (2015) and (b),(d) is the same histogram using our method and (c), (f) is from the method in Matteson and James (2014). From Figures 2.1–2.7 we notice that our method results in more locations being identified as change points in comparison to Preuss et al. (2015). In all models, the number of discoveries at the actual change point locations increases as the sample size increases. It can be seen that for the piecewise i.i.d models (a) and (b), at both the sample sizes 256 and 512, our method performs better than Preuss et al. (2015) in terms of detecting the change points with more accuracy. We also have the comparison with Matteson and James (2014) as this model satisfies their piecewise i.i.d working model assumption. Our method is slightly inferior to their method in results based on model (a) but performs better in results based on model (b) with two change points. This is also the case when some dependence is introduced through a VAR(1) model (c) and the VMA(1) model (d) wherein our method performs better than the competing methods. For the remaining data examples with more dependence, we only present comparisons with Preuss et al. (2015).

The results for the VAR(2) model (e) also indicate better performance of our method with the number of change points estimated at  $\{0.5, 0.75\}$  increasing significantly with increasing series length. For the three dimensional MA(1) model (f) with three change points, the performance is comparable with Preuss et al. (2015) at  $T = 256$ , but it gets better for the larger series length 512. In contrast to earlier examples, the number of change point discoveries is seen to have a significant increase at the location 0.75 but not so much at the other change points. Finally, for the locally stationary bivariate AR(1) model (g), our method has a slightly inferior performance to Preuss et al. (2015) for  $T = 256, 512$ .

Next, in Table 2.1 we present results on the mean number of change points detected per run. Here out of 500 runs of the 7 models discussed above we obtain the average number of change points detected per run. Also reported is the percentage of times the correct number of change points was identified by the 3 competing methods. A detection of a point, say  $\hat{b}$ , is considered correct if  $\hat{b}$  is in a  $N$ -neighborhood (within  $N$  units) of the actual change point location  $b$ . The value of  $N$  here is taken as follows: for any detected point  $\hat{b}$ ,  $N$  is taken as the smallest value in the set  $\{N_1, N_2, \dots, N_n\}$  that led to the addition of  $\hat{b}$  to the set  $P$ ; refer back to Section 2.3.1 for details. For models (a) and (b), we see that our method performs better than Preuss et al. (2015) in terms of identifying the true number of change point locations. For models (c)-(f) with increasing series length, it can be observed that in Preuss et al. (2015), the improvement in detection accuracy i.e % of times the true number of change points is detected is lesser than the proposed method. In model (g) however the method in Preuss et al. (2015) performs better than the competing methods. We notice that the method from Matteson and James (2014) does not perform well for models (b)-(g).

Model		M1		M2		M3	
Sec. 2.8.1.1	T	Mean	% Correct	Mean	% Correct	Mean	% Correct
	256	0.78	75	0.82	76	0.94	94
a (K=1)	512	0.79	76	1.05	85	0.95	95



	1024	0.87	85	1.10	94	0.99	99
b (K=2)	256	0.26	9	0.88	28	1.10	14
	512	0.489	25	1.76	85	1.38	16
	1024	0.791	42	2.08	90	2.95	24
c (K=2))	256	0.22	10	0.48	8	1.72	11
	512	0.38	16	0.683	19	3.85	8
	1024	0.72	34	1.08	41	4.55	6
d (K=2)	256	0.34	10	0.585	19	1.39	9
	512	0.40	15	0.95	34	2.84	6
	1024	0.85	35	1.52	58	4.02	9
e (K=2)	256	0.08	3	0.5	20	1.41	8
	512	0.18	7	0.88	34	2.85	9
	1024	0.45	20	1.45	56	3.90	10
f (K=3)	256	0.38	7	0.45	8	0.74	4
	512	0.72	14	0.802	14	1.18	7
	1024	1.39	36	1.93	47	3.48	20
g (K=1)	256	0.19	15	0.18	14	0.2	5
	512	0.54	48	0.23	21	0.2	5
	1024	0.78	77	0.40	35	0.4	7

Table 2.1: The mean number of change points detected per simulation run of the 7 models in Section 2.8.1.1 for the 3 competing methods M1 (Preuss et al. (2015)), M2 (proposed method) and M3 (Matteson and James (2014)). K denotes the actual number of change points. The % Correct column lists the percentage of runs the correct number or change points was reported by the methods. Results are based on 500 simulation runs of all 7 models.

### 2.8.1.2 Example 2: Empirical Size of the Test

As pointed out by a referee, we investigate the empirical size of the proposed procedure in time series with no change points. We consider 3 stationary models with series length  $T = 128, 256, 512$ . First, a bivariate i.i.d model

$$X_{t,T} = \epsilon_t \quad (2.25)$$

with  $\epsilon_t \sim N(0, I_2)$ . Second, a bivariate AR(1) model

$$X_{t,T} = \begin{bmatrix} 1.4 & 0 \\ 0 & 0.9 \end{bmatrix} X_{t-1} + Z_t \quad (2.26)$$

where  $Z_t$  are i.i.d  $N(0, I_2)$ . Third, tri-variate MA(1) model

$$X_{t,T} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} Z_{t-1} + Z_t \quad (2.27)$$

where  $Z_t$  are i.i.d  $N(0, I_3)$ . The true level of significance (size) was fixed at 1%. In addition, we provide the results for our method but with the  $\infty$ -norm replacing the  $L_2$  norm for the measure of local change given in (2.4). From Table 2.2, a slightly higher estimate of the true size of the proposed test can be seen in the  $L_2$  norm in comparison to the  $\infty$ -norm. This could be viewed as a factor in the higher rate of detection of locations by the proposed method in Section 2.8.1.1 in comparison to the method in Preuss et al. (2015).

	Norm	i.i.d Model	VAR(1)	VMA(1)
T=128	$L_2$ norm	0.02	0.016	0.02
	$\infty$ -norm	0.018	0.02	0.018
T=256	$L_2$ norm	0.012	0.016	0.018
	$\infty$ -norm	0.008	0.014	0.018
T=512	$L_2$ norm	0.008	0.012	0.012
	$\infty$ -norm	0.008	0.008	0.01

Table 2.2: The empirical size of the proposed procedure based on two norm choices ( $L_2$  and  $\infty$ ) used in the test statistic in (2.4). Results are based on 500 simulation runs of the 3 models in Section 2.8.1.2. The true level of significance (size) was fixed at 1%.

### 2.8.1.3 Example 3: Impact of Dimension

We investigate the impact of dimension of the input series on the power of the test for significant change points. For this we consider the following two models with one change point

$$X_{t,T} = I(1 \leq t \leq T/2)\{\epsilon_t + \Theta_1^1 \epsilon_{t-1}\} + I(T/2 < t \leq T)\{\epsilon_t + \Theta_1^2 \epsilon_{t-1}\} \quad (2.28)$$

and

$$X_{t,T} = I(1 \leq t \leq T/2)\{\epsilon_t + \Delta_1^1 \epsilon_{t-1}\} + I(T/2 < t \leq T)\{\epsilon_t + \Delta_1^2 \epsilon_{t-1}\} \quad (2.29)$$

where  $\epsilon_t$  are i.i.d  $N(0, I_p)$ . For  $g \in \{1, 2\}$ , the matrices  $\Theta_1^g$  and  $\Delta_1^g$  are each  $p \times p$  with  $(\Theta_1^1)_{ij} = \beta_1^{|i-j|} I(|i-j| \leq p/2)$ ,  $(\Theta_1^2)_{ij} = (1 + \beta_1)^{|i-j|} I(|i-j| \leq p/2)$  and  $(\Delta_1^1)_{ij} = \beta_2^{|i-j|} I(|i-j| \leq p/2)$ ,  $(\Delta_1^2)_{ij} = (0.5 + \beta_2)^{|i-j|} I(|i-j| \leq p/2)$ . In our simulation we fix  $\beta_1 = 0.5$ ,  $\beta_2 = 0.3$  and repeat the simulation 500 times. We consider dimensions  $p = 4, 6, 8, 10$  and sample sizes  $T = 256, 512, 1024$ . Tables 2.3 and 2.4 present the empirical power estimated over the 500 simulation runs of the two models at various significance levels of the test.

		Model (2.28)				Model (2.29)			
$\alpha$ (size)	T	p=4	p=6	p=8	p=10	p=4	p=6	p=8	p=10
0.01	128	0.3620	0.48	0.5035	0.5487	0.1235	0.2160	0.2998	0.3860
	256	0.4752	0.5766	0.5877	0.6088	0.2464	0.3442	0.4068	0.4550
	512	0.8244	0.9109	0.9205	0.9262	0.4610	0.6042	0.7071	0.7247
	1024	0.9907	1	1	1	0.7745	0.9010	0.9508	0.9734
0.05	128	0.6280	0.70	0.7257	0.7461	0.3005	0.5126	0.5906	0.6587
	256	0.7533	0.8291	0.8548	0.8625	0.4949	0.6123	0.6920	0.7135
	512	0.9591	0.9877	0.9890	0.9850	0.7136	0.8215	0.8940	0.9025
	1024	1	1	1	1	0.9251	0.9412	0.9792	0.9850
0.10	128	0.75	0.8133	0.8465	0.8921	0.5039	0.5994	0.6870	0.77
	256	0.8655	0.9072	0.9126	0.9155	0.6267	0.7038	0.8155	0.8461
	512	0.9856	0.9960	1	1	0.8188	0.9005	0.9370	0.9655
	1024	1	1	1	1	0.9654	0.9875	1	1

Table 2.3: (Based on  $L_2$  norm) The size (first column) and the empirical power (body of the table) of the proposed test against dimension (p) of the input series from 500 simulation runs of the two models given in Section 2.8.1.3.

		Model (2.28)				Model (2.29)			
$\alpha$ (size)	T	p=4	p=6	p=8	p=10	p=4	p=6	p=8	p=10
0.01	128	0.2645	0.3086	0.40	0.4510	0.0875	0.1328	0.1542	0.2168
	256	0.4520	0.5020	0.59	0.5831	0.1950	0.2831	0.3012	0.3235
	512	0.8432	0.8922	0.9320	0.9162	0.4080	0.5911	0.7085	0.7077
	1024	0.99	1	1	1	0.7942	0.9105	0.9508	0.9588
0.05	128	0.5129	0.6042	0.6415	0.6729	0.1891	0.3995	0.4383	0.5527
	256	0.7520	0.7647	0.8133	0.8610	0.4021	0.5322	0.6128	0.6641
	512	0.9321	0.9643	0.99	0.993	0.6560	0.85	0.9053	0.9325
	1024	1	1	1	1	0.9356	0.9421	0.9830	1
0.10	128	0.6628	0.7105	0.76	0.8085	0.3511	0.4874	0.5782	0.6801
	256	0.8435	0.8840	0.9057	0.9252	0.5835	0.7172	0.7638	0.8017
	512	0.98	0.9980	1	0.998	0.7698	0.9260	0.9518	0.9696
	1024	1	1	1	1	0.97	0.99	1	1

Table 2.4: (Based on  $\infty$ -norm) The size (first column) and the empirical power (body of the table) of the proposed test against dimension (p) of the input series from 500 simulation runs of the two models given in Section 2.8.1.3.

We notice from Table 2.3 that for all dimensions, the power increased as the sample size increases as per the asymptotic theory described earlier in Section 2.7. Comparing the results from Tables 2.3,2.4, we notice that for smaller sample sizes the procedure based on the  $L_2$  norm performs better than when replaced by the  $\infty$ -norm. Also at a given level of significance, the power of the test increases as the dimension of the input series increases from 4 to 10. Based on the models described above, this behavior can be attributed to the addition of components that also contain a change point. By adding such components, we notice that our measure of local

change using the  $L_2$  norm on the spectral matrices is such that the signal part (the part that reflects the size of change) increases at the rate  $p^2$  and the noise part at rate  $p$ . Recent methods in Wang and Samworth (2017); Cho and Fryzlewicz (2015) discuss change point methodologies in high-dimensional settings by identifying a sparse subset of components that contribute to the change.

#### 2.8.1.4 Example 4: Confidence Intervals for Change Points

Here we study the construction of confidence intervals (CI) using the technique proposed in Section 2.6. We consider the 7 models from Section 2.8.1.1 with series lengths  $T = 256, 512$ . Out of 500 simulations for both sample sizes, we consider instances where  $\hat{K} = K$ , where  $K$  is the true number of change points in the model. For the choice of  $N$  used for constructing these intervals, we refer back to the procedure described in Section 2.3.1 and select the largest value of  $N_i, i = 1, 2, \dots, n$  for which there is an addition of a point to the set  $P$  in the iterative procedure.

In Table 2.5, the mean and median of the estimate  $\bar{b}$  from (2.10) over 500 runs is listed. The 90% mean lower and upper limits of the confidence intervals along with the coverage probability is also provided. For all models, we notice the increase in coverage probability as series length  $T$  increases from 256 to 512.

Model	Change Pt.	$T$	Mean	Median	Mean 90% CI limits	Coverage Probability
(a)	0.5	256	0.523	0.507	[0.498 , 0.551 ]	86.55
		512	0.510	0.502	[0.491 , 0.521 ]	89.56
(b)	0.3	256	0.335	0.321	[0.282 , 0.397 ]	81.29
		512	0.320	0.307	[0.289 , 0.364 ]	86.25
	0.75	256	0.7656	0.7539	[0.746 , 0.805 ]	83.11
		512	0.760	0.753	[0.744 , 0.779 ]	87.54
(c)	0.5	256	0.550	0.535	[0.452 , 0.569 ]	84.72
		512	0.511	0.505	[0.466 , 0.576 ]	90

	0.75	256	0.781	0.762	[0.743 , 0.796 ]	80.50
		512	0.780	0.762	[0.738 , 0.801 ]	91.04
(d)	0.5	256	0.492	0.484	[0.435 , 0.537 ]	80.67
		512	0.488	0.468	[0.414 , 0.502 ]	85.08
	0.75	256	0.785	0.769	[0.752 , 0.809 ]	84.17
		512	0.761	0.770	[0.748 , 0.802 ]	86.95
(e)	0.5	256	0.496	0.480	[0.429 , 0.527 ]	89.53
		512	0.489	0.475	[0.418 , 0.501 ]	90.10
	0.75	256	0.757	0.757	[0.726 , 0.793 ]	82.60
		512	0.773	0.757	[0.742 , 0.805 ]	87.25
(f)	0.25	256	0.248	0.225	[0.203 , 0.286 ]	82.17
		512	0.245	0.240	[0.224 , 0.278 ]	86.20
	0.5	256	0.488	0.476	[0.429 , 0.574 ]	86.67
		512	0.490	0.484	[0.414 , 0.558 ]	89.11
	0.75	256	0.743	0.734	[0.715 , 0.783 ]	83.34
		512	0.746	0.746	[0.722 , 0.781 ]	89.85
(g)	0.5	256	0.528	0.507	[0.449 , 0.585 ]	91.42
		512	0.515	0.502	[0.482 , 0.588 ]	92.33

Table 2.5: Mean and Median of  $\bar{b}$  from (2.10) over 500 runs. Mean 90% CI limits and estimated coverage probability is listed for the 7 models (a)-(g) from Section 2.8.1.1

## 2.8.2 Application to Seismic Data

Nonstationary time series models have been used to model earthquake data, see Last and Shumway (2008); Ombao et al. (2004) for univariate examples. Strong motion data are measurements over time of ground displacements measured in orthogonal directions. Strong-motion accelerographs are located in major centres of population, near significant faults, or in different

types of building structures and are capable of measuring very strong shaking associated with damaging earthquakes. Ellis and Cakmak (1987); Ellis et al. (1990); Ellis and Cakmak (1991) discuss the advantages of modeling strong motion data in multivariate form and fit VARMA models to the data. We utilize the strong-motion accelerograph data measured in three orthogonal directions, two directions for movement along the ground and one for vertical movement. The data observed over time for the three directions at a given location constitute a tri-variate time series and we aim to capture earthquake occurrences by detecting change points around the earthquake incidents. At all the detected change points, we will also identify the components responsible for seismic activity.

We look at data from the Dannevirke (New Zealand) earthquake of 1975<sup>1</sup> which recorded 5.9 on the Richter scale and had an epicenter about 15km south of Dannevirke. The shocks were felt towards the evening of June 10 and we consider 400 observations from that period of the day. This acceleration data is from the Palmerston North Telephone Exchange station. The three components of the multivariate series are the three orthogonal directions namely N30W, S60W and UP.

We set the level of significance at 5% for the test. For the length of the dyadic neighbourhood  $N$ , we implement the technique in Section 2.3.1 to obtain the change point locations. Our method estimates 4 significant change points given as the dashed lines in Figure 2.8 below. The data reports from Geonet indicate a peak 384.6 m/s/s that can be seen in Figure 7 and list the duration of the shock as 15.9 seconds. Our method rightly detects two change points at time points 287 and 322 which covers the period of peak acceleration for the first two components. For these two points the method also rightly identifies the first two components responsible for the change, see Figure 2.9 for the components contribution plot, where 1, 2, 3 corresponds to components N30W, S60W and UP respectively. In addition it also detects a similar region of peaked acceleration in the first two components between time points 128 and 162 and rightly detect these components as

---

<sup>1</sup>GeoNet(<http://info.geonet.org.nz/display/appdata/Strong-Motion+Data>) is the official source of geological hazard information for New Zealand.



significant towards contributing to seismic activity. The UP component behaves differently with disturbances around time points 225 and 100, but is not deemed as significant by our method. The method in Preuss et al. (2015) detects 6 change points, two of which includes the points 73 and 101 that deems the disturbance in the vertical component as significant. The other detected points are 143, 184, 224 and 300. The method from Matteson and James (2014) estimates change points at 39, 75, 126, 156, 198, 230, 287, 337.

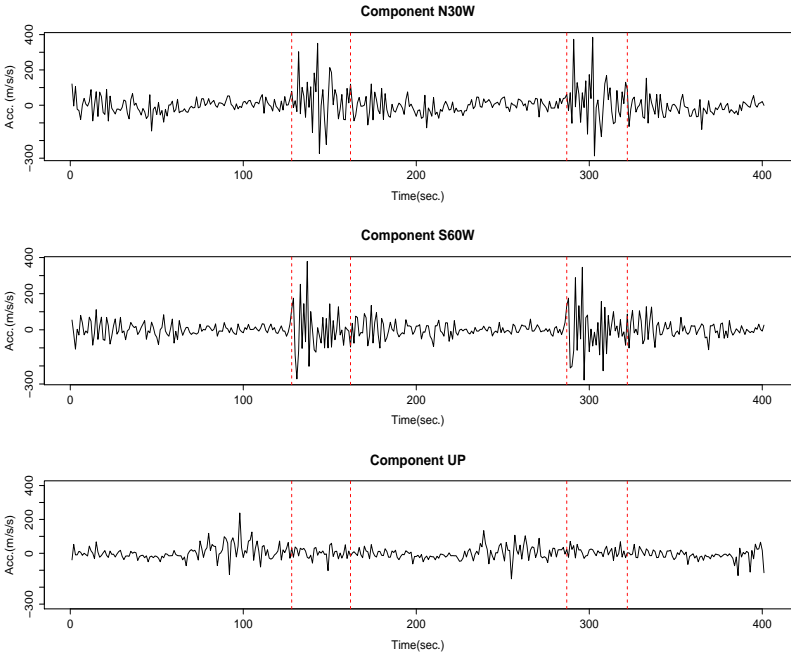


Figure 2.8:  $T = 400$  observations of the Acceleration(m/s/s) readings during the evening of Jun 10, 1975 at Palmerston North Telephone Exchange. Measurements are presented in three orthogonal directions. Dashed lines are change points detected by our method.

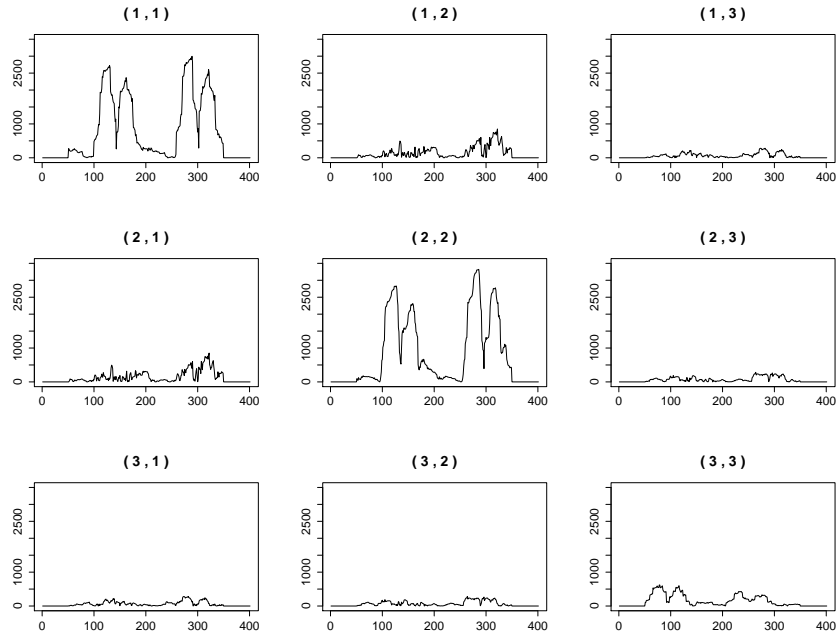


Figure 2.9: Components contribution plot:  $|M_{\hat{b},ij}|$ ,  $i, j = 1, 2, 3$  (y-axis) from (2.8) for time points (x-axis) 51,52,...,350.

### 3. STATIONARY SUBSPACE ANALYSIS OF NONSTATIONARY PROCESSES

In this chapter we begin by describing the SSA model setup and the technique in von Bünau et al. (2009a) for independent random vectors. Then, we present our SSA method from Sundararajan and Pourahmadi (2018b) of finding the stationary subspace of a second-order nonstationary process. Procedures for testing stationarity and determining the dimension of the stationary subspace are presented <sup>1</sup>. Finally, we discuss an application of the proposed method in improving prediction accuracy in neuroeconomic experiments <sup>2</sup>.

#### 3.1 The SSA Setup

Let  $\{X_t\}$  be an observed  $p$ -dimensional nonstationary time series that is a linear combination of  $d$  stationary sources  $Y_t^s \in \mathbb{R}^d$  and  $p - d$  nonstationary sources  $Y_t^n \in \mathbb{R}^{p-d}$ . More precisely,

$$X_t = AY_t = \left[ A_s \mid A_n \right] \begin{bmatrix} Y_t^s \\ Y_t^n \end{bmatrix}, \quad (3.1)$$

where  $A$  is the unknown  $p \times p$  (invertible) mixing matrix,  $A_s$  and  $A_n$  are  $p \times d$  and  $p \times (p - d)$  matrices, respectively. In SSA, the observed vectors are independent and the notion of stationarity is with respect to the first two moments, that is the mean and lag-0 covariance are required to be time-invariant. It is further assumed that no linear transformation of  $Y_t^n$  is stationary. The spaces generated by the columns of  $A_s$  and  $A_n$  are referred to as the *stationary and nonstationary subspaces* of the observed process, respectively.

The objective of SSA is to estimate the demixing matrix  $B = A^{-1}$  so that  $Y_t = BX_t$  is naturally partitioned into its stationary and nonstationary sources. Due to the multiplicative

---

<sup>1</sup>\*Reprinted with permission from “Stationary Subspace Analysis of Nonstationary Processes” by Raanju Ravagavendar Sundararajan and Mohsen Pourahmadi, 2018. Journal of Time Series Analysis, 39 (3), 338-355, Copyright [2018] by John Wiley and Sons.

<sup>2</sup>\*Reprinted with permission from “Reducing Brain Signal Noise in the Prediction of Economic Choices: A Case Study in Neuroeconomics” by Raanju R. Sundararajan, Marco A. Palma and Mohsen Pourahmadi, 2017. Frontiers in Neuroscience, 11, 704, Copyright [2017] by Frontiers.

form of (3.1), the matrix  $B$  is not unique but taken to be an orthogonal matrix. In von Büнау et al. (2009a), the matrix  $B$  is found by dividing the time series data into  $N$  time frames called epochs (segments), and then minimizing as a function of  $B$  the Kullback-Leibler (KL) divergence between Gaussian distributions with given means and covariances across these segments. Let  $\hat{\mu}_i, \hat{\Sigma}_i, i = 1, 2, \dots, N$ , be the sample mean and covariance of the data for the  $i^{th}$  segment, respectively. Consider the  $d \times p$  matrix  $B_1$  as the first  $d$  rows of the  $p \times p$  matrix  $B$ , i.e.

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad (3.2)$$

where  $B_1 \in \mathbb{R}^{d \times p}$  and  $B_2 \in \mathbb{R}^{(p-d) \times p}$ . For a given  $B_1$  it follows that the sample mean and covariance of a candidate stationary source  $Y_t^s = B_1 X_t$  on the  $i^{th}$  segment are

$$\hat{\mu}_i^s = B_1 \hat{\mu}_i \text{ and } \hat{\Sigma}_i^s = B_1 \hat{\Sigma}_i B_1^\top, \quad i = 1, 2, \dots, N. \quad (3.3)$$

The matrix  $B$  is then chosen so that the means and covariances in (3.3) vary the least across all segments. A natural objective function is the sum of the Kullback-Leibler (KL) divergences between the  $N(\hat{\mu}_i^s, \hat{\Sigma}_i^s), i = 1, \dots, N$ , on each epoch and the grand normal distribution  $N(\bar{\mu}^s, \bar{\Sigma}^s)$  where  $\bar{\mu}^s = 1/N \sum_{i=1}^N \hat{\mu}_i^s$  and  $\bar{\Sigma}^s = 1/N \sum_{i=1}^N \hat{\Sigma}_i^s$ :

$$\begin{aligned} L(B) &= \sum_{i=1}^N D_{KL} [ N(\hat{\mu}_i^s, \hat{\Sigma}_i^s) || N(\bar{\mu}^s, \bar{\Sigma}^s) ] \\ &= - \sum_{i=1}^N \left( \log \det \hat{\Sigma}_i^s + (\hat{\mu}_i^s - \bar{\mu}^s)^\top (\hat{\mu}_i^s - \bar{\mu}^s) \right). \end{aligned} \quad (3.4)$$

Then,  $B$  is estimated by minimizing  $L(B)$  using a conjugate gradient descent technique; see von Büнау et al. (2009a), Panknin et al. (2016) for more details. It should be noted that lack of convexity of the objective function may result in many local optima.

Due to the multiplicative form of (3.1), the matrix  $B$  and the latent process are unique up

to scaling, sign and linear transformations within the stationary and nonstationary subspaces. However, even though  $A_s$  and  $A_n$  are not unique, their column spaces denoted by  $\mathcal{C}(A_s)$  and  $\mathcal{C}(A_n)$  are uniquely determined. Interestingly, it is shown in von Büнау et al. (2009a), von Büнау et al. (2009b) that the stationary source in SSA, namely  $\{Y_t^s\}$  is identifiable up to linear transformations provided the number of segments (epochs)  $N$  satisfies,

$$N > \frac{p - d + 1}{2} + 1. \quad (3.5)$$

A negative consequence of the condition (3.5) is that a larger number of nonstationary sources  $p - d$  forces the number of segments  $N$  to be potentially larger, and hence leads to poorer estimates of the means and covariances,  $\hat{\mu}_i^s, \hat{\Sigma}_i^s, i = 1, 2, \dots, N$ , across the segments.

In what follows we refer to the method above as independent-SSA or ISSA, for short. There is also the alternative analytic-SSA (ASSA) proposed in Hara et al. (2012) which reduces solving the challenging non-convex optimization problem in ISSA to a simpler generalized eigenvalue problem, but under the additional assumption that  $Y_t^s$  and  $Y_t^n$  are uncorrelated.

It will be seen later in Section 3.3 that SSA and ASSA are not robust to the choice of  $N$ . Another potential problem in dividing the time series into  $N$  segments is that when the nonstationary sources  $Y_t^n$  is periodically stationary with period  $T/N$ , then the method would not be able to detect the nonstationarity.

### 3.2 Dependent-SSA (DSSA)

In this section we develop an SSA method for finding stationary subspaces of multivariate second-order nonstationary processes satisfying (3.1) using its discrete Fourier transform. In addition to extending ISSA to the dependent data situation, the proposed DSSA method avoids dividing the data into several segments, and is based on a test of stationarity for determining whether the estimated transformation (source) is stationary. A technique to determine the dimension  $d$  of the stationary subspace is given along with some asymptotic properties. See

Sundararajan and Pourahmadi (2018b) for more details.

### 3.2.1 DSSA for Constant Mean Processes

Following the definition in (1.3), we denote discrete Fourier transform (DFT) of a zero-mean  $d$ -variate second-order stationary time series  $Y_t^s$ ,  $1 \leq t \leq T$ , as  $J_Y^s(\omega_k)$  where  $\omega_k = \frac{2\pi}{T}k$ ,  $k = 1, 2, \dots, T$ . Viewing this DFT series as a time series indexed by  $k$ , its lag- $r$  sample autocovariance is the  $p \times p$  complex valued matrix given by

$$\hat{\Gamma}_r^{Y^s} = \frac{1}{T} \sum_{k=1}^T J_{Y^s}(\omega_k) J_{Y^s}(\omega_{k+r})^*, \quad (3.6)$$

where  $J_{Y^s}(\cdot)^*$  denotes the complex conjugate transpose and  $r = 0, 1, \dots$ . Recall from (1.4) that the DFT series is asymptotically uncorrelated at unequal frequencies. The covariance of the DFTs defined above exhibits a very different behavior when  $Y_t^s$  is second-order nonstationary. Lemma A.8 in Jentsch and Subba Rao (2015a) describes the behavior of this quantity when  $Y_t^s$  is a locally stationary vector time series.

Note that  $\hat{\Gamma}_r^{Y^s}$  denotes the lag- $r$  sample DFT autocovariance of  $Y_t^s = B_1 X_t$ . For a given positive integer  $m$ , based on the magnitudes of the first few autocovariances of the DFTs of  $Y_t^s$ , we construct the following objective function as a measure of departure from stationarity:

$$D_Y(B) = \sum_{r=1}^m \left( \| \text{Re}(\hat{\Gamma}_r^{Y^s}(B)) \|_F^2 + \| \text{Im}(\hat{\Gamma}_r^{Y^s}(B)) \|_F^2 \right), \quad (3.7)$$

where for a matrix  $A \in \mathbb{R}^{d \times d}$ ,  $\|A\|_F = \sqrt{\sum_{i,j=1}^d |a_{ij}^2|}$  denotes its Frobenius norm,  $\text{Re}(\cdot)$  and  $\text{Im}(\cdot)$  denote the real and imaginary parts, respectively. The number of lags  $m$  is recommended to be small as larger values could result in loss of power of a stationarity test used in our procedure. For convenience in notation, we drop  $B$  in  $\hat{\Gamma}_r^{Y^s}(B)$  and use  $\hat{\Gamma}_r^{Y^s}$  as the lag- $r$  sample autocovariance of the DFTs.

A solution  $\hat{B}$  is obtained by minimizing  $D_Y(B)$  subject to the assumption of orthogonality,

$BB^\top = I_p$ . Since the objective function in (3.7) is non-convex it could have many local optima. Hence different starting values can be used to obtain a set of solutions. Since the set of orthogonal matrices with determinant 1, namely  $SO(p)$ , constitutes a connected Lie group (see Section 6 of Plumbley (2005)), they can be parametrized as matrix exponentials of skew-symmetric matrices. While restricting the search to  $SO(p)$  results in searching only half the possible permutations of stationary sources, we notice that any permutation and sign change combination of stationary sources is still stationary. In our optimization we focus on the first  $d$  rows of  $B$  and hence any permutation of the  $d$  stationary sources is as good as the other.

The solution can then be obtained using a gradient descent technique by making multiplicative updates. Starting with an orthogonal matrix  $B_0$ , iterative updates of the form  $\hat{B} = e^H B_0$  are made where  $H$  is a skew-symmetric matrix. The exact expressions of the gradient and the update steps needed to solve this optimization problem are provided in the Appendix B.1.

### 3.2.2 DSSA for Time-Varying Mean

In the development of DSSA so far it was assumed that the mean of the observed process is zero. We now discuss the ability of the proposed DSSA in handling nonstationarity when the observed process has a time-varying mean. The key observation is that the components of  $\hat{\Gamma}_r^{Y^s}$  has a non-zero limit or is  $o_p(1)$  depending on whether the mean is time-varying or constant.

Let  $\mu(t/T) \in \mathbb{R}^d$  be the time-varying mean of a  $d$ -variate series  $\{Y_t^s\}$ ,  $1 \leq t \leq T$ , then

$$\begin{aligned} E\left(\sqrt{\frac{2\pi}{T}} J_Y^s(\omega_k)\right) &= E\left(\frac{1}{T} \sum_{t=1}^T Y_t^s \exp(-it\omega_k)\right) \\ &\approx \int_0^1 \mu(u) \exp(-i2\pi uk) du = \sigma_k, \end{aligned} \quad (3.8)$$

where  $\sigma_k$  is the  $d$ -vector of Fourier coefficients of the mean function. It is known that  $|\sigma_{k,i}| \rightarrow 0$  as  $k \rightarrow \infty$  if the  $i^{th}$  component  $\mu_i(\cdot) \in L^2[0,1]$ , where  $\sigma_{k,i}$  is the  $i^{th}$  component of  $\sigma_k$ ,

$i = 1, 2, \dots, d$ . Hence, for any  $a, b = 1, 2, \dots, d$  and lag  $r$ , we have

$$\hat{\Gamma}_{r,a,b}^{Y^s} \approx \sum_{k=1}^T \sigma_{k,a} \sigma_{k+r,b} = O(1). \quad (3.9)$$

On the other hand, if  $\mu(t/T) = \underline{c}$  for some constant vector  $\underline{c}$ , then  $E(J_{Y^s}(\omega_k)) = \underline{0}$  for  $k \neq T\mathbb{Z}$  and we have  $\hat{\Gamma}_{r,a,b}^{Y^s} = o_p(1)$ .

We see that minimizing the left hand side of (3.9) over  $m$  lags results in a stationary subspace where the mean function is nearly constant. However under a time-varying mean, the properties of the test of stationarity described in Section 3.2.4 requires further investigation.

In order to find stationary subspaces of multivariate processes that are both first and second-order nonstationary, a two-stage testing procedure can be employed. For an estimated source  $Y_t^s$ , the first stage tests for constancy of the mean and the second stage, as described in Section 3.2.4, tests for second-order stationarity. Testing for constancy in the mean can be carried out using the fact that under the null hypothesis that  $Y_t^s$  is first and second-order stationary,

$$\begin{pmatrix} \text{Re}(J_{Y^s}(\omega_k)) \\ \text{Im}(J_{Y^s}(\omega_k)) \end{pmatrix} \xrightarrow{D} N \left( \underline{0}, \begin{bmatrix} f_{Y^s}(\omega_k) & \underline{0} \\ \underline{0} & f_{Y^s}(\omega_k) \end{bmatrix} \right) \quad (3.10)$$

where  $f_{Y^s}(\omega_k)$  is the  $d \times d$  spectral matrix of  $Y_t^s$ . Under the alternative hypothesis that the mean is time-varying, we already saw that  $J_{Y^s}(\omega_k) \approx \sqrt{T} \sigma_k$  which supports the use of  $k \ll T$  in the test. Note that to estimate the spectral matrix above it is necessary to either demean the time series or to use  $J_{Y^s}(\omega_k)$  for  $k$  relatively far from zero in the estimation.

### 3.2.3 Identifiability of the Stationary Subspace

Similar to the ISSA, our DSSA method faces the problem of non-uniqueness in the estimated stationary source. In ISSA the condition on the number of epochs given in (3.5) ensures that  $\mathcal{C}(A_s)$  from (3.1) is uniquely determined. In our case, the orthogonality constraint  $BB^T = I_p$  ( or  $B_1 B_1^T = I_d$ ) alone is not sufficient to ensure that  $\mathcal{C}(A_s)$  is uniquely determined. Here we



discuss additional quadratic constraints on  $B_1$  to overcome this problem.

Let  $\mathcal{B} \subset \mathbb{R}^{d \times p}$  be the set of feasible solutions for optimizing (3.7). To have stationarity in the estimated source  $\{Y_t^s\}$ , we search for  $d$  orthonormal vectors so that the magnitude of the DFT covariances of each component of  $\{Y_t^s\}$  is small. More precisely, the orthonormal row vectors  $(b_1, b_2, \dots, b_d)$  of  $B_1 \in \mathbb{R}^{d \times p}$  are chosen to minimize  $|Re(b_j \hat{\Gamma}_r^X b_j^T)| + |Im(b_j \hat{\Gamma}_r^X b_j^T)|$  for  $r = 1, 2, \dots, m$ , where  $m$  is the number of DFT lagged covariances under consideration and  $j = 1, 2, \dots, d$ . Here  $\hat{\Gamma}_r^X$  is the  $p \times p$  lag- $r$  DFT covariance of  $\{X_t\}$  as in (3.6).

Let  $U_r^R$  and  $U_r^I$  be the real and imaginary parts of the  $p \times p$  matrix  $\hat{\Gamma}_r^X$ . We thus have, for  $r = 1, 2, \dots, m$  and  $j = 1, 2, \dots, d$ ,

$$\begin{aligned} |Re(b_j \hat{\Gamma}_r^X b_j^T)| + |Im(b_j \hat{\Gamma}_r^X b_j^T)| &= |b_j U_r^R b_j^T| + |b_j U_r^I b_j^T| \\ &\leq b_j V_r b_j^T \end{aligned}$$

where  $|\cdot|$  denotes the absolute value and the expression for  $V_r$  is given in Section B.2. Now using the above inequality, for any row vector  $b_j$  of  $B_1$ ,  $j = 1, 2, \dots, d$ , in addition to the orthonormality constraint, we impose the following  $m$  quadratic constraints,

$$b_j V_r b_j^T = \delta(T), \quad (3.11)$$

for  $r = 1, 2, \dots, m$ , for a non-negative function  $\delta(T)$  that decays to 0 as  $T \rightarrow \infty$ . The constraint on the  $p$ -vector  $b_j$  in (3.11) forces the upper bound of  $|Re(b_j \hat{\Gamma}_r^X b_j^T)| + |Im(b_j \hat{\Gamma}_r^X b_j^T)|$  to be small.

Next, we provide a lower bound on the number of lags  $m$  required to uniquely identify  $\mathcal{C}(A_s)$ .

**Theorem 3.2.1.** *Let  $\mathcal{B} \subset \mathbb{R}^{d \times p}$  be the space of feasible solutions of (3.7), satisfying the constraints in (3.11) and the orthonormality constraint  $B_1 B_1^T = I_d$  for  $B_1 \in \mathcal{B}$ . Then, in order to uniquely identify  $\mathcal{C}(A_s)$  the number of DFT covariance lags  $m$  must satisfy the inequality*

$$m > p - d. \quad (3.12)$$

*Proof.* See the Appendix A. □

The condition on  $m$  in Theorem 3.2.1 is similar to that on the number of epochs  $N$  in (3.5) in the ISSA setup. Our simulation study reveals the sensitivity of ISSA and ASSA to the choice of  $N$ , and the less sensitive nature of DSSA to the choice of  $m$ .

### 3.2.4 Testing for Stationarity

In this section we provide details of a test of second-order stationarity given in Dwivedi and Subba Rao (2011) and Jentsch and Subba Rao (2015a) using the asymptotic uncorrelatedness of the DFTs of a second-order stationary time series.

Given  $Y_t^s = \hat{B}_1 X_t \in \mathbb{R}^d$ , let  $\hat{f}(\omega)$  be an estimated  $d \times d$  spectral matrix and  $\hat{L}(\omega)$  be the Cholesky factor of its inverse at frequency  $\omega$ . The weighted DFT covariance (Jentsch and Subba Rao, 2015a, Eq. 2.7) is given by

$$\hat{\Gamma}(r, \ell) = \frac{1}{T} \sum_{k=1}^T \hat{L}(\omega_k) J(\omega_k) J(\omega_{k+r})^* \hat{L}(\omega_k)^* \exp(i\ell\omega_k) \quad (3.13)$$

for  $r = 1, 2, \dots, m$  and  $\ell = 0, 1, \dots, n-1$  where  $J(\cdot)^*$ ,  $\hat{L}(\cdot)^*$  denote the complex conjugate transpose. Their test statistic for stationarity is

$$\mathcal{T}(m, n, d) = T \sum_{r=1}^m \left\| W_n^{-1/2} \text{Re}(\hat{K}(r)) \right\|_F^2 + \left\| W_n^{-1/2} \text{Im}(\hat{K}(r)) \right\|_F^2 \quad (3.14)$$

where  $\hat{K}(r) = \left( \text{vech}(\hat{C}(r, 0)), \text{vech}(\hat{C}(r, 1)), \dots, \text{vech}(\hat{C}(r, n-1)) \right)^\top$ ,  $\text{vech}(\hat{C}(r, \ell))$  is the vectorized form of the lower triangular part of the  $d \times d$  matrix  $\hat{\Gamma}(r, \ell)$ , and  $W_n$  is a  $n \frac{d(d+1)}{2} \times n \frac{d(d+1)}{2}$  scaling matrix (Jentsch and Subba Rao, 2015a, Eq. 2.14).

The asymptotic distribution of the test statistic under the null hypothesis that  $\{Y_t^s\}$  is a Gaussian stationary time series is given by

$$\mathcal{T}(m, n, d) \xrightarrow{D} \chi_{mnd(d+1)}^2, \text{ as } T \rightarrow \infty, \quad (3.15)$$

where  $\xrightarrow{D}$  denotes convergence in distribution. The null hypothesis is then rejected if  $\mathcal{T}(m, n, d) > c_{1-\alpha_d}$  where  $c_{1-\alpha_d}$  denotes the  $(1 - \alpha_d)^{th}$  percentile of the  $\chi_{mnd(d+1)}^2$  distribution. Also, under the assumption that  $\{Y_t^s\}$  is locally stationary (alternative hypothesis),  $\mathcal{T}(m, n, d) = O_p(T)$ , where  $m, n$  are assumed to be fixed positive integers.

**Remark 1:** The conditions to be imposed on the lag-window function  $\lambda(\cdot)$  and bandwidth  $b$  while estimating the  $d \times d$  spectral matrix  $\hat{f}(\omega)$  are further discussed in Section 3.2.6.

### 3.2.5 Estimation of $d$ : Dimension of the Stationary Subspace

Here we present a sequential technique for estimating the dimension of the stationary subspace. This is analogous to that in ISSA using a likelihood ratio test as in von Bünau et al. (2009a), Blythe et al. (2012), but without requiring independence of the observations over time.

Given data from the observed  $p$ -variate series  $\{X_t\}$  the sequential search for the dimension  $d$  is as follows. Starting from  $d = 1$ , the demixing matrix  $\hat{B}$  is obtained as a solution by optimizing the discrepancy measure in (3.7). Then the estimated  $Y_t^s = \hat{B}_1 X_t$  is tested for stationarity using the test described in Section 3.2.4. The procedure terminates if there is a rejection of stationarity. Otherwise  $d$  is increased and the steps are repeated. Finally,  $d$  is taken as the largest value for which the test does not reject the null hypothesis of stationarity.

### 3.2.6 Asymptotic Properties of the Estimated $d$

Next, we study some asymptotic properties of the estimated  $d$  using the theoretical properties of the test of stationarity in Section 3.2.4.

For  $\tilde{d} = 1, 2, \dots, p$ , let  $(\mathcal{T}(m, n, \tilde{d}), c(\alpha_{\tilde{d}}))$  denote the test statistic and the critical value at level  $\alpha_{\tilde{d}}$ . Let  $\hat{d}$  be the estimated dimension of the stationary subspace from Section 3.2.5. Note that  $\hat{d}$  can also be written as

$$\hat{d} = \min\{ \tilde{d} = 1, 2, \dots, p : \mathcal{T}(m, n, \tilde{d}) > c(\alpha_{\tilde{d}}) \} - 1. \quad (3.16)$$

We now list the assumptions required for deriving the results.

**Assumption 3.** (A) (i). The observed process  $\{X_{t,T}\}$ ,  $1 \leq t \leq T$ , is a zero-mean locally stationary linear  $p$ -variate time series,

$$X_{t,T} = \sum_{j=0}^{\infty} \Psi_{t,T}(j) \varepsilon_{t-j} \quad (3.17)$$

where  $\varepsilon_t$  are i.i.d  $N_p(0, I_p)$ . Let  $\psi_j : [0, 1] \rightarrow \mathbb{R}^{p \times p}$  be a sequence of functions such that for any  $a, b = 1, 2, \dots, p$ ,

$$|\Psi_{t,T,a,b}(j) - \psi_{j,a,b}(t/T)| \leq \frac{K}{T I(j)},$$

where  $\Psi_{t,T,a,b}(j)$  and  $\psi_{j,a,b}(t/T)$  denote the  $(a, b)^{th}$  component of  $\Psi_{t,T}(j)$  and  $\psi_j(t/T)$  respectively.  $\{I(j)^{-1}\}$  is a monotonically decreasing sequence such that  $\sum_j j^2 I(j)^{-1} < \infty$ .

(ii).  $\sup_u |\psi_j(u)| \leq K/I(j)$  and for  $i = 1, 2, \dots, p$ ,  $\sup_u |\frac{\partial^i}{\partial u^i} \psi_j(u)| \leq K/I(j)$ . Also,  $\sup_u \sum_j |\frac{\partial^2}{\partial u^2} \psi_j(u)| < \infty$ .

(iii). The time-varying  $p \times p$  spectral matrix of  $\{X_{t,T}\}$  is defined as

$$f(u, \omega) = \frac{1}{2\pi} \sum_{r,s=0}^{\infty} \psi_r(u) \psi_s(u)^T e^{-i(r-s)\omega} \quad (3.18)$$

and for  $\omega \in [-\pi, \pi]$ ,  $\int_0^1 f(u, \omega) du$  is non-singular.

(B). The lag-window function  $\lambda : [-1, 1] \rightarrow \mathbb{R}$  used for estimating the spectral matrix in Section 3.2.4 is continuous, symmetric about 0 with bounded first derivative. The bandwidth  $b \in (T^{-1/2}, T^{-1/4})$ .

(C). The number of lags  $m, n$ , used for the test of stationarity in Section 3.2.4 are bounded positive integers.

Assumptions A(i)-A(iii) are fairly standard in the literature of locally stationary linear time series. Next, we state the asymptotic properties of  $\hat{d}$  in estimating the true dimension  $d_0$  of the stationary subspace.

**Theorem 3.2.2.** Suppose that Assumption 1 is satisfied and  $\mathcal{B}$ , the space of feasible solutions

for optimizing (3.7), is defined by the constraints in (3.11) and the orthonormality constraint  $B_1 B_1^T = I_d$  for  $B_1 \in \mathcal{B}$ . Then, as  $T \rightarrow \infty$  we have,

(a). (Consistency)  $P(\hat{d} = d_0) \rightarrow 1$  if  $d_0 = 0$ .

(b).  $P(\hat{d} = d_0) \geq 1 - \sum_{j=1}^{d_0} \alpha_j$  if  $d_0 \in \{1, 2, \dots, p\}$ .

(c).  $P(\hat{d} > d_0) \rightarrow 0$  if  $d_0 \in \{1, 2, \dots, p - 1\}$ .

*Proof.* See the Appendix A. □

Theorem 3.2.2 (a) establishes consistency of the estimator in the absence of stationary sources; an example of this situation is given in our simulation study for model (3.23) in Table 3.5. Theorem 3.2.2 (b) provides a lower bound for the probability of correct selection that depends on the asymptotic levels  $\alpha_1, \alpha_2, \dots, \alpha_{d_0}$ , and (c) shows that with high probability it underestimates the true dimension  $d_0$ .

**Remark 2 :** The presence of multiple tests in the method to detect the dimension requires choice of the rejection levels that balances type I and type II errors. While a small value of rejection levels controls false rejections of the true null and bodes well with the result in Theorem 3.2.2 (b), the rate of false negatives among true alternative hypotheses increases. Implementing a correction to the rejection levels that optimizes both types of errors in our methodology requires further investigation.

### 3.3 Simulation Study

We compare the performance of our DSSA in detecting the actual number of stationary sources to the ISSA in von Büнау et al. (2009a) and ASSA in Hara et al. (2012) using a few simulation studies.

We compare the performance of our DSSA in detecting the actual number of stationary sources to the ISSA in von Büнау et al. (2009a) and ASSA in Hara et al. (2012) using a few simulation studies. For ISSA and ASSA, a sequential likelihood ratio test as in Blythe et al.

(2012) is implemented to select  $d$ . Their sensitivity to the choice of number of epochs  $N$  is also illustrated.

**Study 1:** We begin by considering an i.i.d Gaussian model wherein each component of the stationary sources in  $Y_t^s \in \mathbb{R}^d$  is generated as independent  $N(0, 1)$  random variables. The nonstationary sources are generated using a piecewise i.i.d. Gaussian where for  $i = 1, 2, \dots, p - d$ ,

$$Y_{i,t}^n = \sum_{j=1}^K I(t \in I_j) Y_{i,t}^{(j)} \quad (3.19)$$

is the  $i^{th}$  component of  $Y_t^n$ ,  $I_j$  denotes the  $j^{th}$  interval among  $K$  equal partitions of the interval  $(1, T)$  and  $Y_{i,t}^{(j)} \sim N(0, \sigma_j^2)$  with  $\sigma_j^2$  chosen randomly from the set  $V = \{0.1, 0.5, 1, 2, 3, 4, 4.5\}$ . Choices for the dimension of the input series are  $p \in \{3, 5\}$ . For each  $p$ , we generate models with  $d$  stationary sources and  $p - d$  nonstationary sources where  $d$  is allowed to vary from  $1, \dots, p$ . For every pair of  $(p, d)$ , we generate 100 time series samples of  $Y_t$ , each of length  $T = 400$ , comprising of stationary and nonstationary sources based on the model in (3.19) with  $K = 2$ . A  $p \times p$  orthogonal matrix is randomly generated using the technique from Stewart (1980) and we obtain  $X_t = AY_t$ . Hence for every true pair of  $(p, d)$ , we apply our method to obtain the average p-value from the test of stationarity over 100 simulation runs as in Blythe et al. (2012), and the choice of significance level is avoided. In addition, examining the p-values for the various choices of  $d$  sheds more light on the behavior of the selection procedure rather than just looking at the selected  $d$ . The choice for  $m$ , the number of DFT covariance lags in (3.7), was taken as 5 and 10, but we only present the results from the latter due to similar results for both choices. Similarly, average p-values are obtained using ISSA and the ASSA. The number of epochs,  $N$ , used were 4,  $T/12$  and  $T/16$ , but we only present the best result which was  $N = 4$ . Tables 3.1, 3.2, 3.3, 3.4 provide the performance of our method in detecting the dimension  $d$  of the stationary subspace in comparison to the competing methods.

	True $d$	Estimated $d$		
		1	2	3
DSSA	1	0.4170	0.1012	0
	2	0.4112	0.3542	0
	3	0.5854	0.3446	0.4429
ISSA	1	0.2288	0.1035	0
	2	0.6593	0.7611	0
	3	0.6371	0.7291	0.8159
ASSA	1	0.0002	0	0
	2	0.0001	0	0
	3	0.2193	0.2786	0.8212

Table 3.1: Study 1:  $p = 3$ , Independent Sources: Average p-values of test of stationarity for  $d=1, 2, 3$  for the three competing methods. True values of  $d$  are given in second column.

	True $d$	Estimated $d$				
		1	2	3	4	5
DSSA	1	0.4263	0.0510	0.0231	0	0
	2	0.3825	0.3733	0.0311	0	0
	3	0.5490	0.4631	0.4467	0.1001	0
	4	0.5090	0.5203	0.4586	0.4590	0.0503
	5	0.5270	0.4876	0.5784	0.6179	0.5110
ISSA	1	0.5497	0.0622	0	0	0
	2	0.5644	0.3941	0.0278	0	0
	3	0.6830	0.7244	0.7002	0.1020	0
	4	0.6992	0.8064	0.8852	0.8770	0.2129
	5	0.6908	0.8483	0.8759	0.9400	0.9433
ASSA	1	0.0007	0	0	0	0
	2	0.0041	0.0001	0	0	0
	3	0.0037	0.0045	0	0	0
	4	0.0159	0.0029	0.0122	0.0236	0.0457
	5	0.1190	0.1655	0.3693	0.6380	0.7765

Table 3.2: Study 1:  $p = 5$ , Independent Sources: Average p-values of test of stationarity for  $d=1, 2, 3, 4, 5$  for the three competing methods. True values of  $d$  are given in second column.

From Tables 3.1 and 3.2 we see that in some instances DSSA has inferior performance than ISSA in terms of the estimated average p-values. However, the ASSA technique does not perform very well, as it results too often in a rejection of the test of stationarity.

Next we allow some dependence between the sources by considering the same i.i.d. Gaussian model but the covariance between any two components in  $Y_t$  is randomly chosen from  $(-0.4, 0.4)$ .



The procedure as described earlier is repeated resulting in Tables 3.3 and 3.4 given below. We notice that in a number of instances for the dependent case, DSSA performs better than the ISSA and the ASSA for both dimensions 3 and 5. As an example for  $p = 5$  and true  $d = 1, 3, 4$ , it is seen that DSSA identifies the true dimension  $d$  in a more decisive manner than ISSA. Here again the ASSA results in a large number of incorrect rejections.

	True $d$	Estimated $d$		
		1	2	3
DSSA	1	0.3603	0.1028	0
	2	0.4687	0.4583	0
	3	0.6422	0.4578	0.5470
ISSA	1	0.6393	0.0105	0
	2	0.6581	0.7008	0
	3	0.6351	0.7200	0.8072
ASSA	1	0.0402	0	0
	2	0.0002	0	0
	3	0.2451	0.4542	0.7981

Table 3.3: Study 1:  $p = 3$ , Dependent Sources: Average p-values of test of stationarity for  $d=1, 2, 3$  for the three competing methods. True values of  $d$  are given in second column.

	True $d$	Estimated $d$				
		1	2	3	4	5
DSSA	1	0.5091	0.0500	0.0131	0	0
	2	0.4236	0.3910	0.0511	0	0
	3	0.5289	0.5139	0.5498	0.0601	0
	4	0.5247	0.5127	0.4965	0.5066	0.0700
	5	0.5959	0.4808	0.5316	0.6437	0.5178
ISSA	1	0.5091	0.1966	0.0083	0	0
	2	0.6260	0.5980	0.0178	0	0
	3	0.6368	0.7117	0.6742	0.1720	0
	4	0.6292	0.8165	0.9052	0.8945	0.3824
	5	0.6635	0.8356	0.9121	0.9535	0.9511
ASSA	1	0.0004	0	0	0	0
	2	0.0001	0.0001	0	0	0
	3	0.0003	0	0	0	0
	4	0.0001	0.0002	0	0	0
	5	0.1357	0.1553	0.3425	0.6563	0.8382

Table 3.4: Study 1:  $p = 5$ , Dependent Sources: Average p-values of test of stationarity for  $d=1, 2, 3, 4, 5$  for the three competing methods. True values of  $d$  are given in second column.

**Study 2:** In this study we compare the number of times the correct dimension  $d$  is selected for the two extreme cases of multivariate stationary ( $d = p$ ) and nonstationary ( $d = 0$ ) processes and other classes in between ( $d = 1, 2, \dots, p - 1$ ). We simulate from the following three classes of models:

**Bivariate Stationary VAR(2) Model:**

$$X_t = \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} X_{t-1} + \begin{bmatrix} 0.5 & 0.2 \\ 0.1 & 0.6 \end{bmatrix} X_{t-2} + \varepsilon_t \quad (3.20)$$

where  $\varepsilon_t$  are i.i.d.  $N(0, I_2)$ .

**Nonstationary Unit Root VAR(1) Model:** Consider bivariate autoregressive models

$$X_t = \Phi X_{t-1} + \varepsilon_t \quad (3.21)$$

where  $\varepsilon_t$  are i.i.d.  $N(0, \Lambda)$  and the polynomial  $I - \Phi L$  has a unit root corresponding to the following cases; see Ahn and Reinsel (1990) and Li et al. (2001):

$$\Phi_1 = \begin{bmatrix} 0.6 & 1 \\ 0.12 & 0.7 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 25 & 5.4 \\ 5.4 & 9 \end{bmatrix}. \quad (3.22)$$

Note that the matrices  $I - \Phi_j$ ,  $j = 1, 2, 3$ , have rank 1 and the eigenvalues of  $\Phi_1, \Phi_2, \Phi_3$  are  $(1, 0.3)$ ,  $(1, 0.5)$  and  $(1, 0.8)$ , respectively. We refer to these models as  $(\Phi_i, \Lambda)$ ,  $i = 1, 2, 3$  in Table (3.5), where  $\Lambda$  is as above or the identity matrix  $I_2$ . Hence the stationary subspace has dimension 1.

**Trivariate Nonstationary MA(1) Model:**

$$X_t = \Theta_t \varepsilon_{t-1} + \varepsilon_t, \quad (3.23)$$

where where  $\varepsilon_t$  are i.i.d.  $N(0, I_3)$  and  $\Theta_t$  is given by

$$\begin{aligned} \Theta_t = & I\left(\frac{t}{T} \leq 0.25\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} + I\left(0.25 < \frac{t}{T} \leq 0.5\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \\ & + I\left(0.5 < \frac{t}{T} \leq 0.75\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + I\left(\frac{t}{T} > 0.75\right) \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}. \end{aligned}$$

We generate the time series  $\{X_t\}$  of lengths  $T = 400, 800$ , for each of the above models and compare the performances of DSSA, ISSA and ASSA by the number of times the true dimension was detected out of 100 replications. The significance level for the tests used was fixed at 0.01. From Table 3.5 we see that DSSA more accurately detects the true dimension than ISSA and ASSA in all the cases. The sensitivity to the choice of  $N$ , the number of epochs in ISSA and ASSA, can be seen and the lack of sensitivity to the choice of  $m$ , the number of DFT covariance lags in DSSA, is witnessed. For the DSSA, it can also be seen that the number of times the true dimension was detected increases with increasing sample size. The results from the unit root autoregressive models in (3.22) indicate the poor performance of ISSA and ASSA as opposed to the heteroscedastic type models given in (3.19). This could be due to their requirement of independent observations for the likelihood ratio test. In contrast, the DSSA performs much better in these models mainly because it is developed to work for the general class of multivariate second-order nonstationary time series without imposing any independence assumption. Comparing the percentages in the third and fourth rows of Table 3.5 also reveals the negative impact of the dependence in the noise  $\varepsilon_t$  on the correct selection of the true  $d$ .

Model	$T$	DSSA		ISSA			ASSA		
		$m = 5$	$m = 10$	$N = 5$	$N = 10$	$N = 30$	$N = 5$	$N = 10$	$N = 30$
VAR(2)	400	96	94	81	62	23	56	29	9
	800	97	98	78	61	11	48	40	87
$(\Phi_1, \Lambda)$	400	74	73	1	7	5	7	1	0
	800	80	82	11	13	2	11	12	1
$(\Phi_1, I_2)$	400	81	83	3	8	3	4	12	3
	800	89	89	10	4	10	2	5	3
$(\Phi_2, I_2)$	400	76	76	4	2	11	4	10	0
	800	80	79	6	3	10	6	6	2
$(\Phi_3, I_2)$	400	82	82	4	9	2	7	6	10
	800	85	86	5	10	4	8	1	2
VMA(1)	400	99	100	55	40	19	94	95	92
	800	99	99	98	66	10	99	98	98

Table 3.5: Study 2: Number of times the true dimension  $d$  was detected by DSSA, ISSA and ASSA for the models (3.20), (3.22) and (3.23).

### 3.4 A Case Study in Neuroeconomics

A total of 181 right-handed students participated in a food snack choice decision experiment conducted in the Texas A & M Human Behavior Laboratory. The sample consisted of about 50 % females and 50% males. The subjects were presented with 10 food choice task questions (10 trials). Each choice consisted of two food products, product A and product B. The two products within each alternative had the same features relative to brand, price, packaging and flavor. The only difference between each pair of products was that one of them had fewer calories, making it

a healthier choice. For example: original strawberry Jello –70 calories– (vs) sugar-free strawberry Jello –10 calories. See Table 3.6 below for the list of product choices. The focus of this paper is reducing EEG noise to improve the prediction of which of the two food snacks participants would choose, irrespective of the product's identity. Subjects were asked to fast for three hours prior to the experiment, and received a compensation fee of \$20 in exchange for their participation. In order to incentivize and make the food choice task real, one of the 10 tasks was randomly selected to be binding and participants had to eat the food snack before being paid and leaving the laboratory. The displayed picture of each item was the actual photo available for purchase in Walmart's website; however, the participants were not aware that the products were purchased in Walmart. The above case study and the description of the analysis provided in the next few sections in this chapter are adapted from Sundararajan et al. (2017).

The experimental design proceeds as follow. At the beginning of the experiment, a blank slide with a fixation point in the middle of the computer screen was presented for 2 seconds. Then, for each food choice task, the actual product images were presented in the following screen for 8 seconds. A separate decision slide asked participants which of the two food snacks they prefer to eat. After each decision, an inter-stimulus slide was presented for 0.75 seconds. The order of the products was randomized across trials in the experimental design; however, all subjects completed the task in the same order.

### **3.4.1 Data Acquisition**

The participant was fitted with a proper size EEG headset (B-Alert X10, Advanced Brain Monitoring, Inc.) with 9 electrodes to record brain activity from the pre-frontal (F3, F4, FZ), central (C3, C4, CZ), and parietal (P3, P4, POZ) cortices and a linked mastoid reference. An electrode impedance test was performed to ensure proper conductivity of the electrodes. The impedance level threshold was  $20k\Omega$ . An EEG calibration procedure was implemented before the data collection. The EEG calibration incorporated choice tasks (unrelated to the study), psychomotor, and auditory psychomotor vigilance tasks. The EEG data was collected at a sampling

Choice	Product A	Product B
1	Low calorie Jell-O (10 cal.)	Original Jell-O (70 cal.)
2	Oven baked Lays Chips (120 cal.)	Classic Lays Chips (160 cal.)
3	No sugar Dole peaches (25 cal.)	Original Dole peaches (70 cal.)
4	Light Yoplait (90 cal.)	Original Yoplait (150 cal.)
5	Fat free Pringles (70 cal.)	Original Pringles (150 cal.)
6	Sugar free Snack Pack (70 cal.)	Original Snack Pack (110 cal.)
7	Reduced fat Sargento cheese (50 cal.)	Original Sargento cheese (80 cal.)
8	Non-fat Greek yogurt (120 cal.)	Original Greek yogurt (150 cal.)
9	Reduced fat Cheez-It (130 cal.)	Original Cheez-It (150 cal.)
10	Diet Lipton tea (0 cal.)	Traditional Lipton tea (100 cal.)

Table 3.6: Food Snack Choice Questions

rate of 256Hz. The experiment was presented using the iMotions software platform.

### 3.4.2 Data Analysis

For any given food choice task, say product A vs product B, we gathered the 9-dimensional EEG signal from the 9 electrodes from the start of the stimuli when the product images are shown to 2.5 seconds after the start. On the digital signal scale, this constitutes 640 observations ( $2.5 \times 256$ ). More precisely, for each subject  $j = 1, 2, \dots, 181$ , the data comprises of 640 observations across time.

Given the raw 9-dimensional EEG time series obtained in this case study, we proceed according to the following algorithm to obtain the prediction results:

The Prediction Algorithm:

Step 1: Filter the raw 9-dimensional signal using a 0.5 Hz high-pass and 45 Hz low-pass filter.

Denote the filtered series as  $\{X_{t,j}\}$  where  $j = 1, 2, \dots, 181$  and  $t = 1, 2, \dots, 640$ .

Step 2: Pre-whiten  $\{X_{t,j}\}$ . For convenience in notation, we denote  $X_{t,j}$  as the band-pass filtered signal that has been pre-whitened.

Step 3: Noise reduction: Apply SSA to  $\{X_{t,j}\}$  to obtain  $\{Y_{t,j}\}$  (Section 3.2 or 3.2).

Step 4: Feature Selection and Prediction Models (Section 3.4.3)

Step 5: Assessing Prediction Performance (Section 3.4.3.1)

In Step 2 we pre-whiten  $\{X_{t,j}\}$  before further analysis by computing the  $9 \times 9$  sample covariance matrix  $S_j$  and then transform the data to  $(S_j)^{-0.5}X_{t,j}$ . This standardization reduces the cross-sectional correlation in  $\{X_{t,j}\}$ .

#### 3.4.2.1 *Noise Reduction via SSA*

It is common to treat data like  $\{X_{t,j}\}$  as a nonstationary time series; see Ombao et al. (2005), Park et al. (2014) for examples. The words noise and nonstationarity are used interchangeably because in our setup the nonstationary sources contribute to parts of the signal that are unrelated to the food choice task. Hence eliminating nonstationarity reduces noise in the brain signal. As an illustration, we make a plot of the 9-dimensional EEG signal  $X_{t,j}$  (before noise reduction) in Figure 3.1. In Figure 3.2, we then plot a 3-dimensional stationary subspace process obtained after application of SSA. The presence of nonstationarity (noise) in  $X_{t,j}$  was confirmed by carrying out formal tests of stationarity; Jentsch and Subba Rao (2015a). Hence we resort to the SSA technique for removing this nonstationarity from the signal and this is described in Sections 3.1, 3.2.



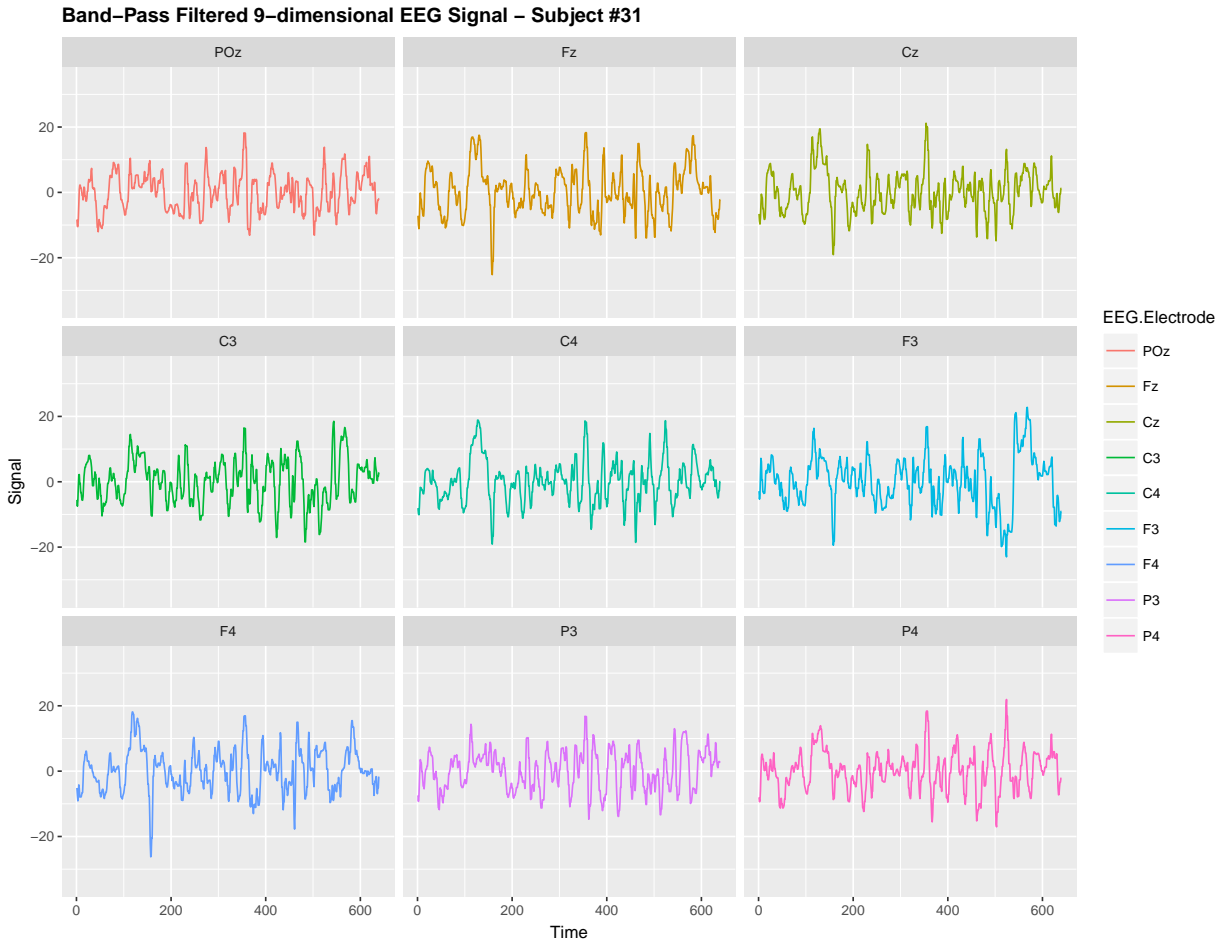


Figure 3.1: Band-Pass filtered 9-dimensional EEG signal  $\{X_{t,j} : t = 1, 2, \dots, 640\}$  (before noise reduction) gathered from subject # 31 while responding to food-choice question number 9.



Figure 3.2: 3-dimensional stationary subspace process  $\{Y_{t,j} : t = 1, 2, \dots, 640\}$  (after noise reduction) gathered from subject #31 while responding to food-choice question number 9.

As a pre-processing technique to reduce noise, we apply DSSA and ISSA described in Sections 3.1, 3.2 to obtain a  $d$  dimensional stationary subspace process where  $d < 9$ , denoted by  $\{Y_{t,j}\}$ . Since the actual dimension  $d$  is unknown, we present the results for  $d = 4, 5, 6, 7, 8$ . We also applied the sequential technique in Sundararajan and Pourahmadi (2018b) to detect  $d$  for each subject and each food choice task. Here we obtained a mode of  $d = 8$  as an estimate of the dimension of the stationary subspace.

### 3.4.3 The Prediction Models

We discuss three prediction models based on logistic regression with different derived features. The aim of the prediction models discussed below is to fit a model to predict product choice (A or B) based on the input signal. While building prediction models  $M_1$  and  $M_2$ , only Step 2 of the algorithm is used, for prediction model  $M_3$  both Steps 2 and 3 are needed. Note that model  $M_2$  assumes that  $\{X_{t,j}\}$  is stationary whereas model  $M_3$  assumes that  $\{X_{t,j}\}$  is nonstationary and applies SSA before extracting features and estimating the prediction model.

**Model  $M_1$ :** A standard model similar to Telpaz et al. (2015) is based on the importance of the pre-frontal EEG channels in explaining choice behavior in individuals. Following their aggregation technique to reduce the noise when computing preference scores for products, we take average of the signals from the 3 pre-frontal channels (F3, F4, FZ) over the 2.5 seconds. The signal here is a 3-dimensional band-pass filtered signal that was pre-whitened (Step 1). The average is taken per subject per food choice question (say product A vs product B). For subject  $j$ ,  $j = 1, 2, \dots, 181$ , this average denoted by the scalar  $\bar{X}_j$  is used as a feature in the following logistic regression model:

$$P\left(c_{j,AB} = 1 \mid \bar{X}_j\right) = \frac{\exp(a_0 \bar{X}_j)}{1 + \exp(a_0 \bar{X}_j)}, \quad (3.24)$$

for  $j = 1, 2, \dots, 181$ . In the model above we have denoted 1 for product A and 0 for product B and the model predicts the class (0 or 1) based on the derived feature  $\bar{X}_j$ .

**Model  $M_2$ :** In this approach, to distinguish between the two classes denoted as 1 for product A and 0 for product B, we take  $\{X_{t,j}\}$  the pre-whitened 9-dimensional band-pass filtered signal. We then focus on the covariance structure of  $X_{t,j}$  for each of the two classes (0 and 1). The aim is to derive features that bring out the differences between the two classes based on the covariance structure of the signal. This is achieved by computing the average spectral density

matrices for the two classes over the Fourier frequencies:

$$\bar{g}^i(\omega_k) = \frac{1}{n_i} \sum_{j \in \text{Class } i} g_j(\omega_k), \quad i = 0, 1, \quad (3.25)$$

where  $g_j(\omega_k)$  is the estimated  $9 \times 9$  spectral matrix for subject  $j$  using observations  $\{X_{t,j}\}$ ,  $n_i$  for  $i = 0, 1$  is the number of subjects in the two classes and  $\omega_k = \frac{2\pi k}{640}$ ,  $k = 1, 2, \dots, 640$ , are the fundamental Fourier frequencies. The spectral matrix was estimated using a Daniell kernel with smoothing window length 25 (approximately  $\sqrt{640}$ ); see Example 10.4.1 in Brockwell and Davis (1991).

In order to train the classifier, for every subject  $j \in \{1, 2, \dots, 181\}$ , a distance vector  $p_{j,AB} = (p_{0,j,AB}, p_{1,j,AB})$  is computed where

$$p_{i,j,AB} = \frac{1}{640} \sum_{k=1}^{640} \|g_j(\omega_k) - \bar{g}^i(\omega_k)\|_F^2 \quad i = 0, 1.$$

and  $\|\cdot\|_F$  is the Frobenius norm of a matrix. It measures the distance to the center of each of the two classes and serves as our two-dimensional feature vector used in constructing the following logistic regression model (prediction model):

$$P\left(c_{j,AB} = 1 \mid p_{j,AB}\right) = \frac{\exp(\alpha_0 p_{0,j,AB} + \alpha_1 p_{1,j,AB})}{1 + \exp(\alpha_0 p_{0,j,AB} + \alpha_1 p_{1,j,AB})}, \quad (3.26)$$

for  $j = 1, 2, \dots, 181$  and  $c_{j,AB}$  is the class indicator (1 for product A or 0 for product B) for subject  $j$ .

**Model  $M_3$ :** Here we apply Step 2 on the raw 9-dimensional EEG signal to obtain  $\{X_{t,j}\}$ . Then we obtain on the  $d$ -variate stationary subspace processes,  $\{Y_{t,j}\}$ , using DSSA/ISSA (Step 2). Similar to the approach in model  $M_2$ , we aim to capture the differences between the two classes based on the covariance structure of the signal. Unlike model  $M_2$ , we apply DSSA and ISSA described in Sections 3.1, 3.2 to obtain a  $d$ -dimensional stationary subspace process where

$d < 9$ , denoted by  $\{Y_{t,j}\}$ . Features to be fed into the prediction model will be based on  $\{Y_{t,j}\}$  as opposed to model  $M_2$  wherein  $\{X_{t,j}\}$  was used. Then, proceeding as in model  $M_2$ , we compute the average spectral density matrices for the two classes over the Fourier frequencies:

$$\bar{f}^i(\omega_k) = \frac{1}{n_i} \sum_{j \in \text{Class } i} f_j(\omega_k), \quad i = 0, 1, \quad (3.27)$$

where  $f_j(\omega_k)$  is the estimated  $d \times d$  spectral matrix for subject  $j$  using observations  $\{Y_{t,j}\}$ ,  $n_i$  for  $i = 0, 1$  is the number of subjects in the two classes and  $\omega_k = \frac{2\pi k}{640}$ ,  $k = 1, 2, \dots, 640$  are the fundamental Fourier frequencies. The spectral matrix was estimated using a Daniell kernel with smoothing window length 25 (approximately  $\sqrt{640}$ ).

In order to train the classifier, for every subject  $j \in \{1, 2, \dots, 181\}$ , a distance vector  $d_{j,AB} = (d_{0,j,AB}, d_{1,j,AB})$  is computed where

$$d_{i,j,AB} = \frac{1}{640} \sum_{k=1}^{640} \|f_j(\omega_k) - \bar{f}^i(\omega_k)\|_F^2 \quad i = 0, 1.$$

and  $\|\cdot\|_F$  is the Frobenius norm of a matrix. It measures the distance to the center of each of the two classes and serves as our two-dimensional feature vector used in constructing the following logistic regression model (prediction model):

$$P\left(C_{j,AB} = 1 \mid d_{j,AB}\right) = \frac{\exp(\beta_0 d_{0,j,AB} + \beta_1 d_{1,j,AB})}{1 + \exp(\beta_0 d_{0,j,AB} + \beta_1 d_{1,j,AB})}, \quad (3.28)$$

for  $j = 1, 2, \dots, 181$  and  $C_{j,AB}$  is the class indicator (1 for product A or 0 for product B) for subject  $j$ .

### 3.4.3.1 Prediction Performance

We assess the performance by computing the overall prediction accuracy and the average sensitivity and specificity. Using the confusion matrix given in Table 3.7, we compute two prediction

accuracy measures given by

$$A_1 = \frac{C_A + C_B}{T_A + T_B}, \quad A_2 = \frac{\frac{C_A}{T_A} + \frac{C_B}{T_B}}{2}, \quad (3.29)$$

where  $A_1$  is the overall prediction accuracy of the model and  $A_2$ , in a binary classification context, is the average of sensitivity (true positive rate) and specificity (true negative rate) of the prediction models.

		Prediction		Total
		Product A	Product B	
Actual	Product A	$C_A$	$I_B$	$T_A$
	Product B	$I_A$	$C_B$	$T_B$

Table 3.7: Confusion matrix

Finally, we present an estimate of the AUC: area under the ROC curve (LeDell et al. (2015)) for the 10 food choice questions for each of the 3 models and this measure is denoted as  $A_3$ . The ROC curve plots the true positive rate against the false positive rate and is a useful measure of model performance. The area under the ROC curve (known as AUC) varies between 0-100% with a value of 50% as baseline (uninformative classifier).

In Table 3.8, we shuffle the class labels randomly and fit the prediction models and assess the performance measures. The shuffling of labels is done 500 times, each time fitting the prediction models, and the average performance measure over the 500 runs across the 10 food choice questions is presented. This enables us to identify a baseline for the 3 performance measures (70% for performance measure  $A_1$  and 50% for performance measures  $A_2$  and  $A_3$ ).

Model	Overall Accuracy - $A_1$	Avg. of Sensitivity and Specificity - $A_2$	AUC - $A_3$
$M_1$	69.54	48.97	53.26
$M_2$	69.05	51.81	53.25
$M_3$ - DSSA	69.12	51.34	54.20
$M_3$ - ISSA	69.51	51	52.08

Table 3.8: Prediction performance of the 3 models with shuffled labels: the average of the 3 performance measures  $A_1$ ,  $A_2$  and  $A_3$  (AUC) taken across the 10 food choice questions for the three competing models  $M_1$ ,  $M_2$  and  $M_3$ . For model  $M_3$  the choice of  $d$  is taken as 8.

These accuracy rates are computed using a 10-fold cross-validation technique where the data is randomly divided into 10 nearly equal parts. Each part is removed, in turn, while the remaining data are used to fit the prediction models  $M_1$ ,  $M_2$ ,  $M_3$  and the predictions are carried out for the left out part. More precisely, the computed accuracy rates are the out-of sample estimates wherein for any given pair of products A and B, the prediction model is fit based on roughly 90% of the subjects and the predictions are carried out for the remaining subjects.

The overall accuracy rate ( $A_1$ ) for models  $M_1$  and  $M_2$  computed and plotted in Figure 3.3 shows that it varies between 69-72% for both models. Next, we look at the performance measure  $A_2$  as an average of the sensitivity and specificity of models  $M_1$  and  $M_2$ . We notice from Figure 3.4 that both methods perform poorly with accuracy rates around 50%. Note that as opposed to averaging over the signal across the 3 channels in model (3.24), we also assessed the performance of the logistic regression models fitted individually with each of the pre-frontal channels. We obtained rates (not presented here) similar to that seen in Figures 3.3 and 3.4 in terms of overall prediction accuracy and average of sensitivity and specificity.

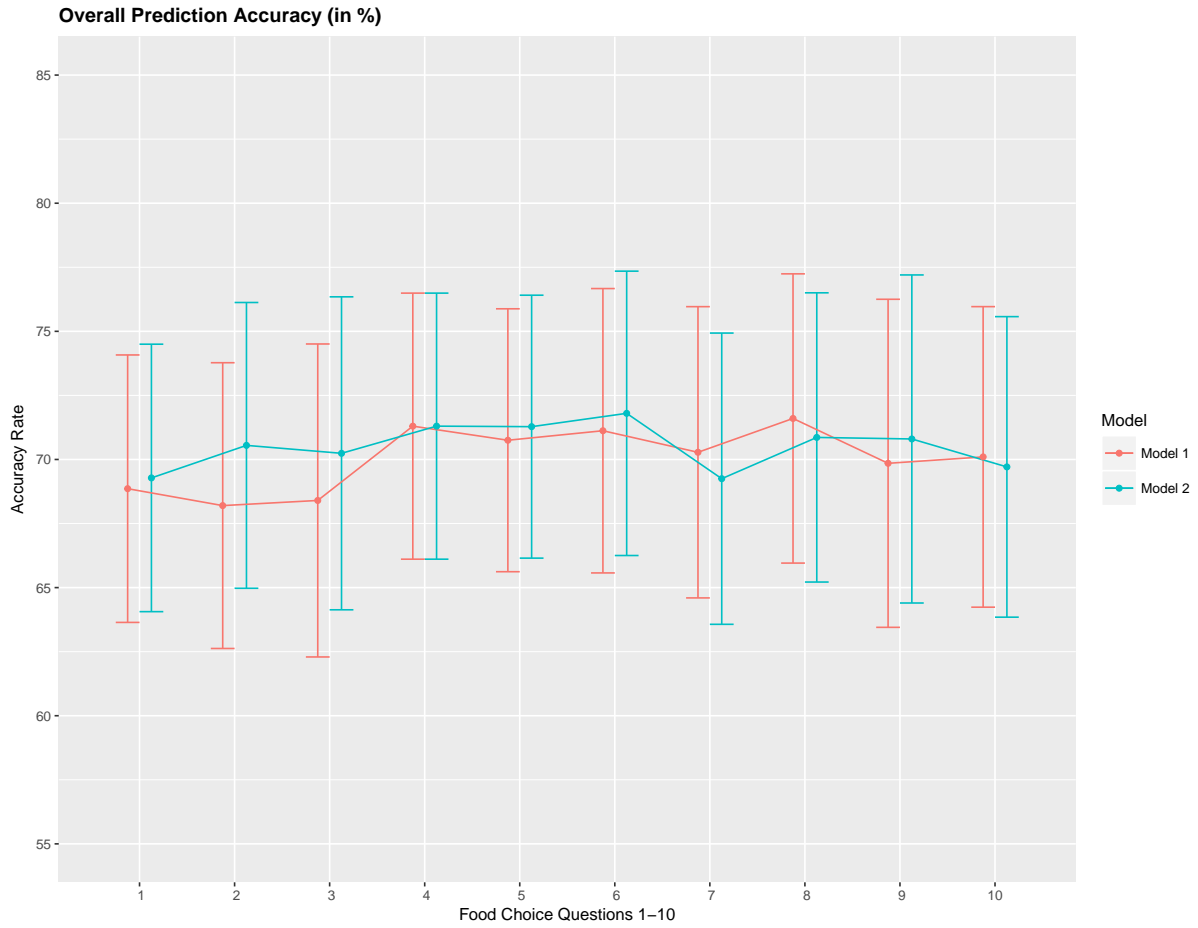


Figure 3.3: Overall prediction accuracy rate ( $A_1$ ), in %, based on a 10-fold cross-validation for the 10 food choice tasks for the two models  $M_1$  and  $M_2$ . Approximate 95% confidence intervals included for each accuracy estimate.



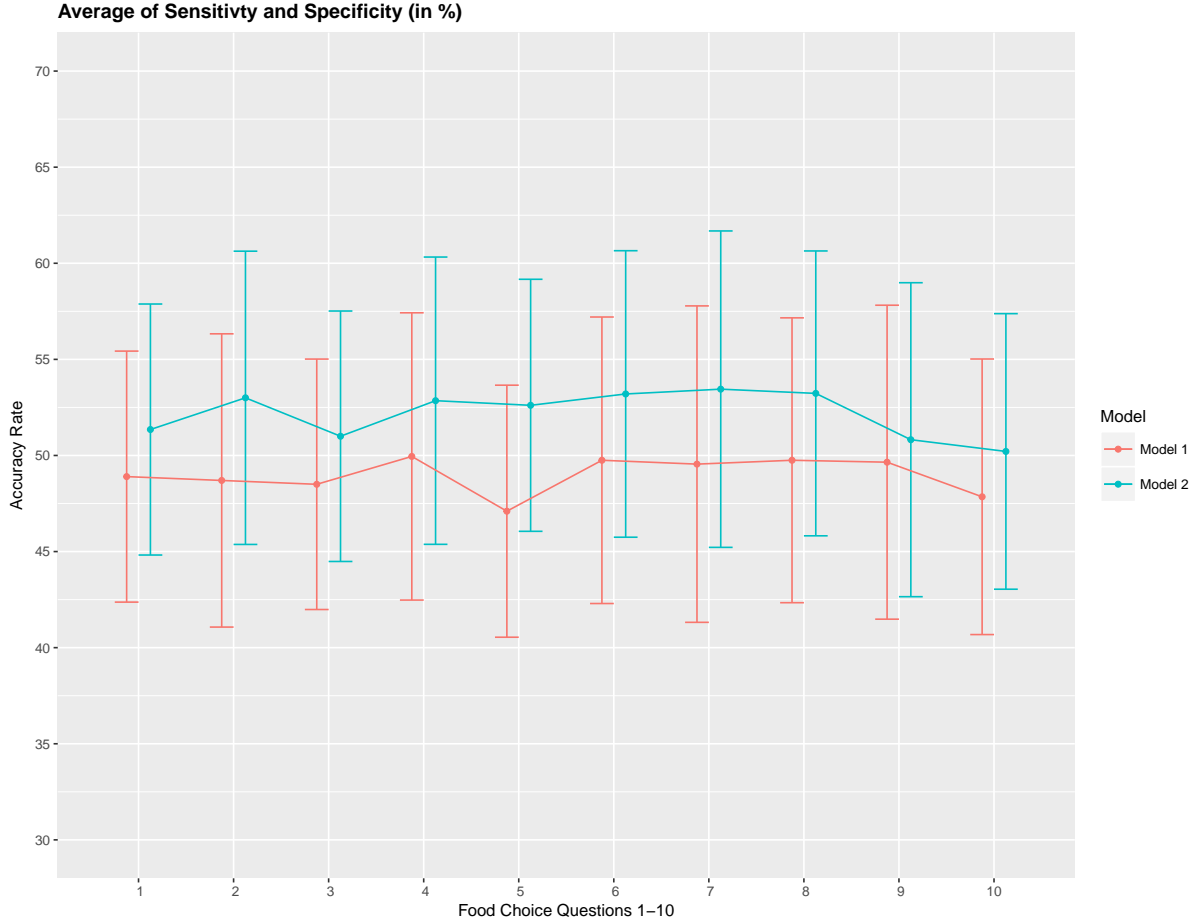


Figure 3.4: Prediction accuracy rate as average of sensitivity and specificity ( $A_2$ ), in %, based on a 10-fold cross-validation for the 10 food choice questions for the two models  $M_1$  and  $M_2$ . Approximate 95% confidence intervals included for each accuracy estimate.

Next, we study the overall prediction accuracy ( $A_1$ ) after applying the pre-processing techniques DSSA and ISSA and removing the nonstationarity (noise) in the EEG signal, and fitting the prediction model (3.28) (model  $M_3$ ). Since the actual dimension of the stationary subspace is unknown, in Table 3.9 we present the results for dimensions  $d = 4, 5, 6, 7, 8$ , which show that DSSA performs better than ISSA in most cases. The average overall accuracy rate based on the 10 food choice tasks for each value of  $d$  is given in Figure 3.5. It is seen that the 10-fold cross-validation accuracy rate is around 80% for each of the 10 tasks when the dimension  $d = 8$ .

This rate is roughly 10% more than the accuracy rate from Figure 3.3 wherein no SSA-type pre-processing technique is applied. We also notice that as the dimension of the stationary subspace  $d$  increases, the accuracy rate also increases. This phenomenon was also observed in von Bünau et al. (2010) and confirms the improvements in prediction accuracy when there are fewer nonstationary sources (noise) in model (3.1). The DSSA/ISSA turns out to be a very useful tool for reducing the noise (nonstationarity) in the EEG signal.

$d$		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
4	DSSA	<b>72.80</b>	<b>73.65</b>	74.22	71.53	<b>74.00</b>	73.83	<b>73.48</b>	<b>71.38</b>	<b>74.50</b>	72.97
	ISSA	69.60	71.20	75.10	71.40	66.50	76.20	70.18	68.45	72.37	73.13
5	DSSA	73.50	73.21	<b>75.76</b>	<b>73.00</b>	68.30	<b>77.85</b>	73.55	72.40	<b>74.60</b>	<b>78.75</b>
	ISSA	74.00	75.48	71.80	71.80	75.10	75.10	75.20	74.62	72.40	70.70
6	DSSA	<b>74.65</b>	75.00	<b>78.46</b>	72.85	77.35	<b>75.10</b>	75.70	74.52	<b>77.45</b>	<b>78.61</b>
	ISSA	71.94	77.30	74.00	79.62	78.50	73.50	75.70	76.80	74.00	74.25
7	DSSA	76.12	<b>80.27</b>	<b>78.56</b>	79.15	79.20	76.15	79.65	80.11	<b>79.12</b>	<b>80.50</b>
	ISSA	75.70	74.00	76.20	81.80	80.12	79.80	78.50	80.80	77.90	76.80
8	DSSA	<b>78.98</b>	<b>81.35</b>	<b>81.52</b>	<b>84.12</b>	<b>80.24</b>	79.11	<b>82.30</b>	80.58	80.13	<b>81.94</b>
	ISSA	76.90	75.10	80.16	79.60	75.10	79.00	79.00	84.75	82.00	75.68

Table 3.9: 10-fold cross-validation overall prediction accuracy (in %) for the 10 questions Q1-Q10 corresponding to  $d=4, 5, 6, 7, 8$  for DSSA and ISSA (model  $M_3$ ). Significant results (instances of at least a 1% improvement in DSSA) are highlighted in bold.

We then assess the performance measure  $A_2$  which is an average of the sensitivity and specificity for model  $M_3$ . We set the dimension of the stationary subspace at  $d=8$ . Figure 3.6 shows that DSSA performs slightly better than ISSA in most cases. More importantly, we note that in comparison to Figure 3.5, DSSA has roughly a 20% increase in the performance measure  $A_2$ .

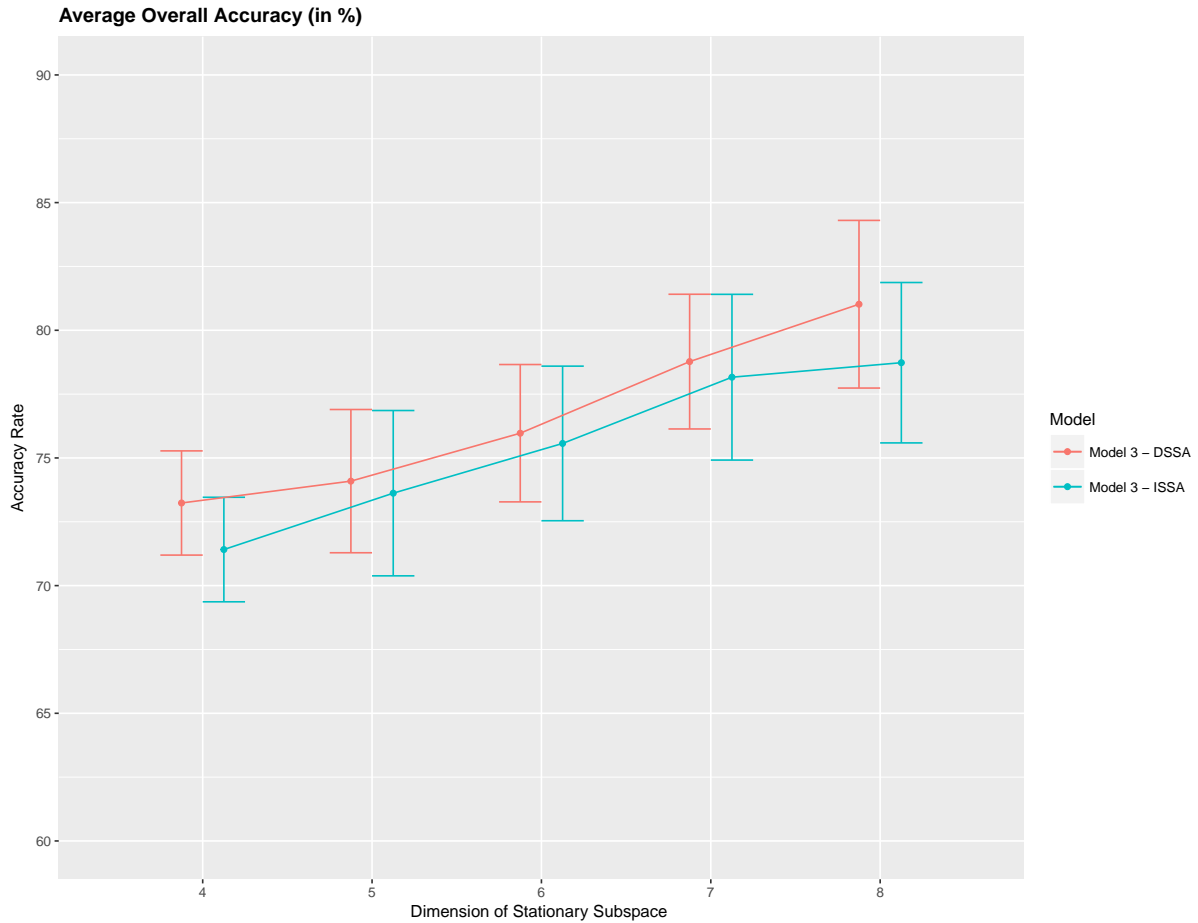


Figure 3.5: Average 10-fold cross-validation overall accuracy rate (in %) for the 10 food choice questions (y-axis) versus dimension of the stationary subspace (x-axis). Approximate 95% confidence intervals included for each accuracy estimate.

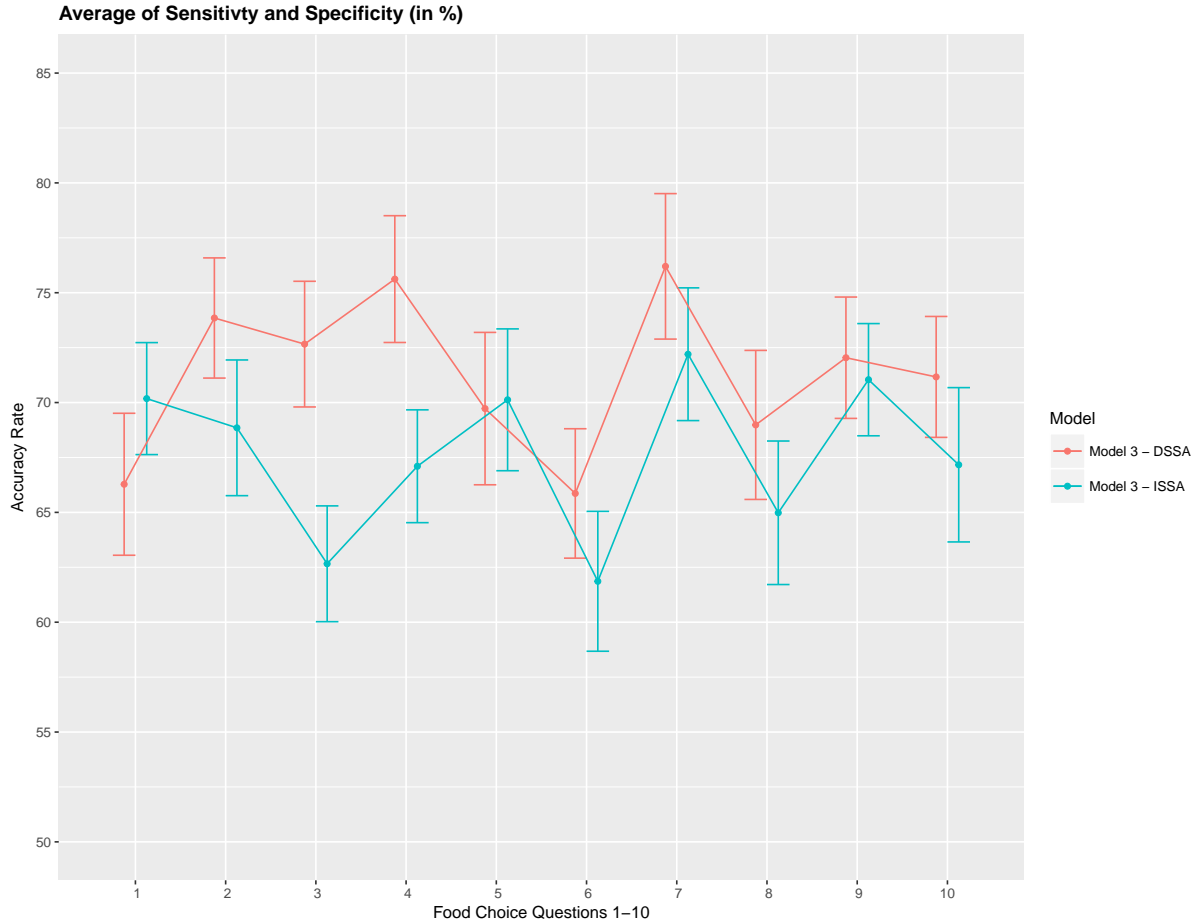


Figure 3.6: Prediction accuracy rate: average of sensitivity and specificity ( $A_2$ ), in %, based on a 10-fold cross-validation for the 10 food choice questions. Model  $M_3$  was used with  $d = 8$ . Approximate 95% confidence intervals included for each accuracy estimate.

Finally, we present a cross-validation estimate of the AUC for the 3 competing models in Figure 3.7. We again notice roughly a 20% increase when using DSSA/ISSA (Model  $M_3$ ) as a noise reduction technique before constructing the prediction model.

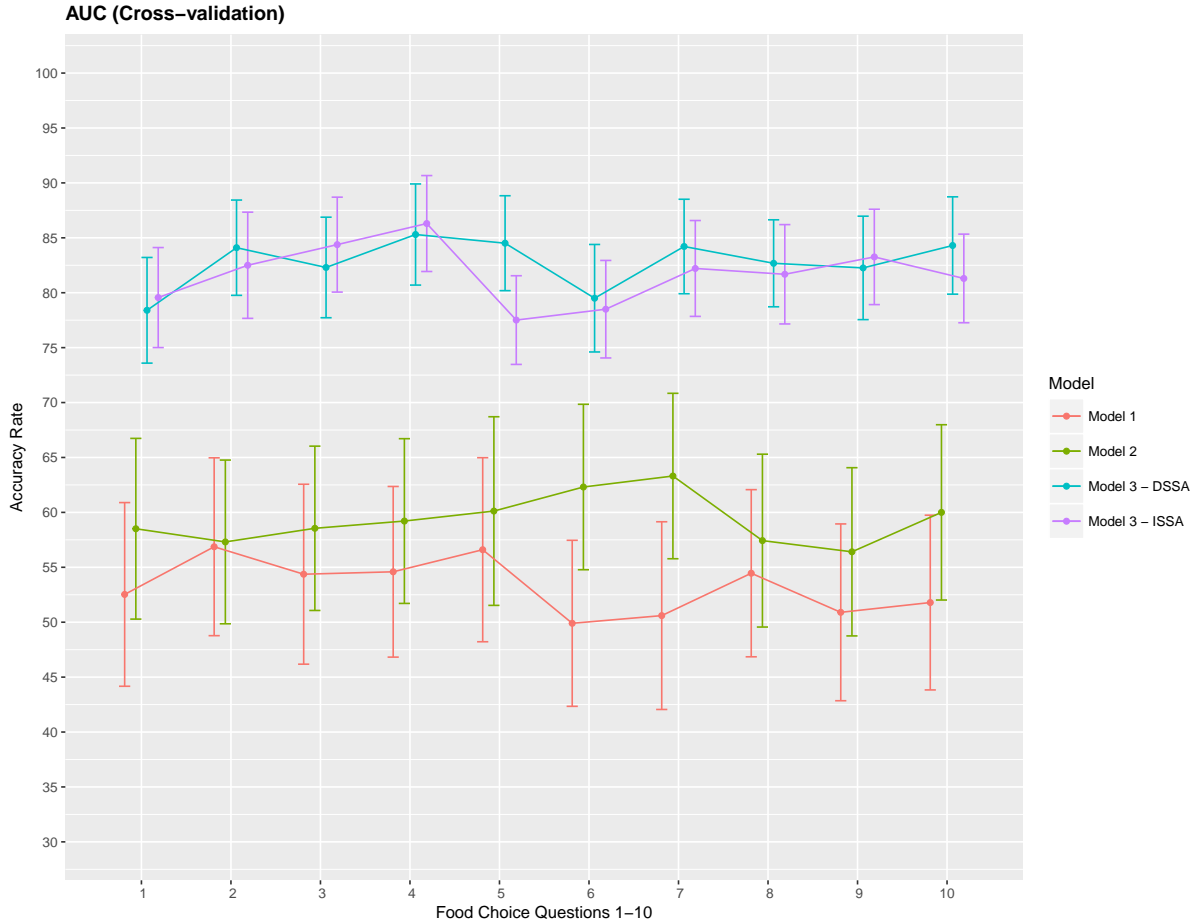


Figure 3.7: Cross-validation estimate of the AUC in % (Area under the ROC curve) for the 3 models  $M_1$ ,  $M_2$  and  $M_3$ . Approximate 95% confidence intervals included for each accuracy estimate. For model  $M_3$  we take  $d = 8$ .

The average of the 3 performance measures  $A_1$ ,  $A_2$  and  $A_3$ (AUC) taken across the 10 food choice questions for the three competing models  $M_1$ ,  $M_2$  and  $M_3$  is reported in Table 3.10. We note that for models  $M_1$  and  $M_2$  the overall prediction accuracy ( $A_1$ ) is roughly 70% which is treated as a baseline for this measure. However, the performance measures  $A_2$  and  $A_3$  (AUC) are only around 50% which suggests a poor performance. In contrast for model  $M_3$ , overall accuracy rate increased by roughly 10%, the measure  $A_2$  is higher by around 20% and measure  $A_3$  (AUC) is significantly higher (increase of roughly 30%) than models  $M_1$  and  $M_2$ .

Model	Overall Accuracy - $A_1$	Avg. of Sensitivity and Specificity - $A_2$	AUC - $A_3$
$M_1$	70.04	48.97	53.26
$M_2$	70.52	52.17	59.31
$M_3$ - DSSA	81.02	71.23	82.75
$M_3$ - ISSA	78.73	67.61	81.72

Table 3.10: The average of the 3 performance measures  $A_1$ ,  $A_2$  and  $A_3$  (AUC) taken across the 10 food choice questions for the three competing models  $M_1$ ,  $M_2$  and  $M_3$ . For model  $M_3$  the choice of  $d$  is taken as 8.

## 4. SUMMARY AND CONCLUSIONS

In Chapter 2, we have proposed a new method to detect multiple change points in multivariate nonstationary time series. Our method focuses on identifying changes in the second-order properties of the process and estimates change point locations by quantifying changes in the spectral matrices through a squared Euclidean norm. The numerical, methodological and theoretical advantages of our method has been presented. The procedure is completely nonparametric with only a few tuning parameters to be chosen.

The presence of multiple tests in our method potentially causes an increase in the overall error rate and the critical values in our procedure are obtained pointwise. Obtaining critical values to control the overall error requires further theoretical investigation. Also of interest would be controlling the uniform type-I error rate of our test statistic while testing for existence of at least one change point in the time series. The construction of simultaneous confidence intervals for multiple change points can also be studied with Bonferroni-type corrections on the confidence levels. Change point detection for infinite variance processes is another interesting problem that is currently being pursued.

In Chapter 3, we have proposed an SSA method for finding stationary linear transformations of multivariate nonstationary processes. As opposed to much of the SSA literature that consider nonstationarity based on the mean and lag-0 covariance, we take up second-order nonstationary processes that includes the entire covariance structure. The DSSA method exploits the near-uncorrelatedness of the DFT of a second-order stationary time series. A sequential testing procedure is proposed to select the dimension of the stationary subspace and its asymptotic properties are discussed. A case study in neuroeconomics is presented wherein we illustrate the usefulness of our method in improving prediction accuracy in neuroeconomic experiments.

This work also leads to a number of interesting research problems that are worth studying:

1. The theoretical properties established for the estimator of dimension  $d$  implicitly assume

that our procedure obtains an estimate  $\hat{B}_1$  where  $\hat{B}_1 \in \mathcal{C}(\hat{A}_s)$  and  $\mathcal{C}(\hat{A}_s) \rightarrow \mathcal{C}(A_s)$  as  $T \rightarrow \infty$ . This requires establishing consistency in the estimated column space  $\mathcal{C}(\hat{A}_s)$  which is postponed for future work.

2. The result in Theorem 3.2.2 (b) establishes a lower bound on the probability of correct dimension selection. Constructing estimators of the true dimension that improve this lower bound is of practical and theoretical importance.
3. Under the setup in (3.1), the problem of finding the stationary subspace can be examined using a factor model approach with the dimension of the subspace now treated as the number of factors.
4. ISSA analogue of DSSA for the broad class of second-order nonstationary processes would be of great interest.



## REFERENCES

- Adak, S. (1998). Time dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association* 93(444), 1488–1501.
- Ahn, S. K. and G. C. Reinsel (1990). Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* 85(411), 813–823.
- Aue, A., S. Hormann, L. Horvath, and M. Reimherr (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* 37, 4046–4087.
- Azimmohseni, M., A. R. Soltani, and M. Khalafi (2015). Simulation of real discrete time gaussian multivariate stationary processes with given spectral densities. *Journal of Time Series Analysis* (1).
- Blythe, D. A. J., P. von Bunau, F. C. Meinecke, and K. R. Muller (2012, April). Feature extraction for change-point detection using stationary subspace analysis. *IEEE Transactions on Neural Networks and Learning Systems* 23(4), 631–643.
- Boksem, M. A. and A. Smidts (2015). Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success. *Journal of Marketing Research* 52(4), 482–492.
- Brillinger, D. (2001). *Time Series*. Society for Industrial and Applied Mathematics.
- Brockwell, P. J. and R. A. Davis (1991). In *Time Series: Theory and Methods*. Springer-Verlag: Berlin.
- Cardinali, A. and G. P. Nason (2011). Costationarity of locally stationary time series. *Journal of Time Series Econometrics* 2(2).
- Chen, J. and A. K. Gupta (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association* 92(438), pp. 739–747.
- Cho, H. and P. Fryzlewicz (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B*

- (*Statistical Methodology*) 77(2), 475–507.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* 25(1), 1–37.
- Dahlhaus, R. (2000, 12). A likelihood approximation for locally stationary processes. *Ann. Statist.* 28(6), 1762–1794.
- Dahlhaus, R. (2012). Locally stationary processes. In S. S. R. Tata Subba Rao and C. Rao (Eds.), *Time Series Analysis: Methods and Applications*, Volume 30 of *Handbook of Statistics*, pp. 351 – 413. Elsevier.
- Dette, H. and E. Paparoditis (2009). Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. *Journal of the Royal Statistical Society: Series B* 71(4), 831–857.
- Dwivedi, Y. and S. Subba Rao (2011). A test for second-order stationarity of a time series based on the discrete fourier transform. *Journal of Time Series Analysis* 32(1), 68–91.
- Eichler, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *Journal of Multivariate Analysis* 99(5), 968–1009.
- Ellis, G. and A. Cakmak (1991). Time series modelling of strong ground motion from multiple event earthquakes. *Soil Dynamics and Earthquake Engineering* 10(1), 42 – 54.
- Ellis, G., R. DeVeaux, and A. Cakmak (1990). Multivariate time series modelling of strong motion accelerograms recorded in mexico and taiwan. *Soil Dynamics and Earthquake Engineering* 9(5), 218 – 235.
- Ellis, G. W. and A. S. Cakmak (1987, August). Modelling earthquake ground motions in seismically active regions using parametric time series methods. Technical Report NCEER-87-0014, Department of Civil Engineering, Princeton University, Princeton, New Jersey.
- Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55(2), 251–276.
- Granger, C. (1981). Some properties of time series data and their use in econometric model

- specification. *Journal of Econometrics* 16(1), 121 – 130.
- Hara, S., Y. Kawahara, T. Washio, P. von Büna, T. Tokunaga, and K. Yumoto (2012). Separation of stationary and non-stationary sources with a generalized eigenvalue problem. *Neural Networks* 33, 7 – 20.
- Helgason, H., V. Pipiras, and P. Abry (2011, May). Fast and exact synthesis of stationary multivariate gaussian time series using circulant embedding. *Signal Process.* 91(5), 1123–1133.
- Hu, B. and W. Schiehlen (1997). On the simulation of stochastic processes by spectral representation. *Probabilistic Engineering Mechanics* 12(2), 105–113.
- Jentsch, C. and S. Subba Rao (2015a). A test for second order stationarity of a multivariate time series. *Journal of Econometrics* 185(1), 124 – 161.
- Jentsch, C. and S. Subba Rao (2015b). A test for second order stationarity of a multivariate time series. *Journal of Econometrics* 185(1), 124 – 161.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59(6), 1551–1580.
- Kaplan, A. Y., A. A. Fingelkurts, A. A. Fingelkurts, S. V. Borisov, and B. S. Darkhovsky (2005). Nonstationary nature of the brain activity as revealed by eeg/meg: Methodological, practical and conceptual challenges. *Signal Processing* 85(11), 2190 – 2212. *Neuronal Coordination in the Brain: A Signal Processing Perspective*.
- Khushaba, R. N., L. Greenacre, S. Kodagoda, J. Louviere, S. Burke, and G. Dissanayake (2012). Choice modeling and the brain: A study on the electroencephalogram (eeg) of preferences. *Expert Systems with Applications* 39(16), 12378–12388.
- Khushaba, R. N., C. Wise, S. Kodagoda, J. Louviere, B. E. Kahn, and C. Townsend (2013). Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (eeg) and eye tracking. *Expert Systems with Applications* 40(9), 3803–3812.
- Killick, R., C. F. Nam, J. Aston, and I. Eckley (2012). *changeoint.info: The changepoint*

repository.

- Kirch, C., B. Muhsal, and H. Ombao (2015). Detection of changes in multivariate time series with application to eeg data. *Journal of the American Statistical Association* 110(511), 1197–1216.
- Kreiss, J.-P. (1992). Bootstrap procedures for  $ar(\infty)$  — processes. In K.-H. Jöckel, G. Rothe, and W. Sendler (Eds.), *Bootstrapping and Related Techniques: Proceedings of an International Conference, Held in Trier, FRG, June 4–8, 1990*, pp. 107–113. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Last, M. and R. Shumway (2008). Detecting abrupt changes in a piecewise locally stationary time series. *Journal of Multivariate Analysis* 99(2), 191 – 214.
- LeDell, E., M. Petersen, and M. van der Laan (2015). Computationally efficient confidence intervals for cross-validated area under the roc curve estimates. *Electron. J. Statist.* 9(1), 1583–1607.
- Lemm, S., B. Blankertz, T. Dickhaus, and K.-R. Müller (2011). Introduction to machine learning for brain imaging. *NeuroImage* 56(2), 387 – 399. Multivariate Decoding and Brain Reading.
- Li, W. K., S. Ling, and H. Wong (2001). Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity. *Biometrika* 88(4), 1135–1152.
- Matteson, D. S. and N. A. James (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association* 109(505), 334–345.
- Messer, M., M. Kirchner, J. Schiemann, J. Roeper, R. Neininger, and G. Schneider (2014, 12). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *Ann. Appl. Stat.* 8(4), 2027–2067.
- Meyer, M. and J.-P. Kreiss (2015). On the vector autoregressive sieve bootstrap. *Journal of Time Series Analysis* 36(3), 377–397.
- Muggeo, V. and G. Adelfio (2011). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics* 27(2), 161–166.

- Nieto, F. H., D. Pena, and D. Saboya (2016). Common seasonality in multivariate time series. *Statistica Sinica* 26(4).
- Ombao, H., J. Heo, and D. Stoffer (2004). Online analysis of seismic signals. In D. R. Brillinger, E. A. Robinson, and F. P. Schoenberg (Eds.), *Time Series Analysis and Applications to Geophysical Systems*, Volume 45 of *The IMA Volumes in Mathematics and its Applications*, pp. 53–71. Springer New York.
- Ombao, H., R. von Sachs, and W. Guo (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association* 100(470), 519–531.
- Panknin, D., P. von Büнау, M. Kawanabe, F. C. Meinecke, and K.-R. Müller (2016). Higher order stationary subspace analysis. *Journal of Physics: Conference Series* 699(1), 012021.
- Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scandinavian Journal of Statistics* 27(1), 143–176.
- Park, T., I. A. Eckley, and H. C. Ombao (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Transactions on Signal Processing* 62(20), 5240–5250.
- Patton, A., D. N. Politis, and H. White (2009). Correction to “automatic block-length selection for the dependent bootstrap” by d. politis and h. white. *Econometric Reviews* 28(4), 372–375.
- Peña, D. and P. Poncela (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference* 136(4), 1237 – 1257.
- Percival, D. B. and W. L. B. Constantine (2006). Exact simulation of gaussian time series from nonparametric spectral estimates with application to bootstrapping. *Statistics and Computing* 16(1), 25–35.
- Petersen, K. B. and M. S. Pedersen (2012, nov). The matrix cookbook. Version 20121115.
- Picard, F., S. Robin, M. Lavielle, C. Vaisse, and J.-J. Daudin (2005). A statistical approach for array cgh data analysis. *Bioinformatics* 6(27), 1–14.
- Plumbley, M. D. (2005). Geometrical methods for non-negative ica: Manifolds, lie groups and

- toral subalgebras. *Neurocomputing* 67, 161 – 197. Geometrical Methods in Neural Networks and Learning Geometrical Methods in Neural Networks and Learning.
- Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association* 89(428), 1303–1313.
- Politis, D. N. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23, 53–70.
- Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*. John Wiley & Sons Inc.
- Preuss, P., R. Puchstein, and H. Dette (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association* 110(510), 654–668.
- Ravaja, N., O. Somervuori, and M. Salminen (2013). Predicting purchase decision: The role of hemispheric asymmetry over the frontal cortex. *Journal of Neuroscience, Psychology, and Economics* 6(1), 1.
- Stewart, G. W. (1980). The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Journal on Numerical Analysis* 17(3), 403–409.
- Sun, T. C. and M. Chaika (1997). On simulation of a gaussian stationary process. *Journal of Time Series Analysis* 18(1), 79–93.
- Sundararajan, R. R., M. A. Palma, and M. Pourahmadi (2017). Reducing brain signal noise in the prediction of economic choices: A case study in neuroeconomics. *Frontiers in Neuroscience* 11, 704.
- Sundararajan, R. R. and M. Pourahmadi (2018a). Nonparametric change point detection in multivariate piecewise stationary time series. *minor revision - Journal of Nonparametric Statistics*.
- Sundararajan, R. R. and M. Pourahmadi (2018b). Stationary subspace analysis of nonstationary processes. *Journal of Time Series Analysis* 39(3), 338–355.
- Telpaz, A., R. Webb, and D. J. Levy (2015). Using eeg to predict consumers' future choices. *Journal of Marketing Research* 52(4), 511–529.

- Venkatraman, V., A. Dimoka, P. A. Pavlou, K. Vo, W. Hampton, B. Bollinger, H. E. Hershfield, M. Ishihara, and R. S. Winer (2015). Predicting advertising success beyond traditional measures: New insights from neurophysiological methods and market response modeling. *Journal of Marketing Research* 52(4), 436–452.
- von Büнау, P., F. C. Meinecke, S. Scholler, and K. R. Müller (2010, Aug). Finding stationary brain sources in eeg data. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 2810–2813.
- von Büнау, P., F. C. Meinecke, F. C. Király, and K.-R. Müller (2009a, Nov). Finding stationary subspaces in multivariate time series. *Phys. Rev. Lett.* 103, 214101.
- von Büнау, P., F. C. Meinecke, F. C. Király, and K.-R. Müller (2009b). Finding stationary subspaces in multivariate time series: Supplementary material. *EPAPS Document*, <http://www.aip.org/pubservs/epaps.html> no.E-PRLTA-103-014948.
- Wang, T. and R. J. Samworth (2017). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, n/a–n/a.
- Webb, R., P. W. Glimcher, I. Levy, S. C. Lazzaro, and R. B. Rutledge (2013). Neural random utility and measured value. *Available at SSRN*.
- Yau, C. Y. and Z. Zhao (2015). Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78(4), 895–916.

## APPENDIX A

### TECHNICAL PROOFS

**Proof of Theorem 2.7.1.** Under Assumptions 1,2 we can apply Theorem 3.5 of Eichler (2008) to establish asymptotic normality. It can also be seen that the required Assumption 3.2 of Eichler (2008) is also satisfied by our test statistic  $\hat{D}(b)$ . We now obtain the mean and variance of the test statistic.

Firstly, we write  $2\pi Nh^{1/2}\hat{D}(b)$  as

$$\begin{aligned} & 2\pi Nh^{1/2} \int_{-\pi}^{\pi} \|\text{vec}(\hat{f}_L(b, \omega) - \hat{f}_R(b, \omega))\|^2 d\omega \\ &= 2\pi Nh^{1/2} \int_{-\pi}^{\pi} \|\text{vec}(\hat{f}_{11}^{(b)}(\omega) - \hat{f}_{22}^{(b)}(b, \omega))\|^2 d\omega \\ &= 2\pi Nh^{1/2} \int_{-\pi}^{\pi} \|\text{vec}\left(\frac{1}{N} \sum_{k=-\lfloor(N-1)/2\rfloor}^{\lfloor N/2\rfloor} K_h(\omega - \omega_k) \{ I_{11}^{(b)}(\omega_k) - I_{22}^{(b)}(\omega_k) \}\right)\|^2 d\omega \end{aligned}$$

where  $I_{11}^{(b)}(\omega_k)$  and  $I_{22}^{(b)}(\omega_k)$  are the  $p \times p$  periodogram matrices at frequency  $\omega_k$  of the series  $\{Y_{1,t}^{(b)}\}$  and  $\{Y_{2,t}^{(b)}\}$  respectively.

Now, for the expectation we have

$$\begin{aligned} E\left(2\pi Nh^{1/2}\hat{D}(b)\right) &= \frac{2\pi h^{1/2}}{N} \int_{-\pi}^{\pi} \sum_{i,j=1}^p \sum_{k_1, k_2} K_h(\omega - \omega_{k_1}) K_h(\omega - \omega_{k_2}) \times \\ & E\left[\left(I_{11,ij}^{(b)}(\omega_{k_1}) - I_{22,ij}^{(b)}(\omega_{k_1})\right) \overline{\left(I_{11,ij}^{(b)}(\omega_{k_2}) - I_{22,ij}^{(b)}(\omega_{k_2})\right)}\right] d\omega \end{aligned}$$

where for  $m_1, m_2 = 1, 2$ ,  $I_{m_1 m_2, ij}$  denotes the  $(i, j)^{th}$  entry of the  $p \times p$  matrix  $I_{m_1 m_2}$ . Under  $H_0$  it can be seen that the expectation term on the right hand side vanishes if  $\omega_{k_1} \neq \omega_{k_2}$ . So the



expected value becomes

$$\begin{aligned}
E\left(2\pi N h^{1/2} \hat{D}(b)\right) &= \frac{2\pi}{N h^{3/2}} \int_{-\pi}^{\pi} \sum_{i,j=1}^p \sum_k K^2\left(\frac{\omega - \omega_k}{h}\right) E\left[ I_{11,ij}^{(b)} \overline{I_{11,ij}^{(b)}} + \right. \\
&\quad \left. I_{22,ij}^{(b)} \overline{I_{22,ij}^{(b)}} - I_{11,ij}^{(b)} \overline{I_{22,ij}^{(b)}} - I_{22,ij}^{(b)} \overline{I_{11,ij}^{(b)}} \right] d\omega \\
&= \frac{2\pi}{N h^{3/2}} \int_{-\pi}^{\pi} \sum_{i,j=1}^p \sum_k K^2\left(\frac{\omega - \omega_k}{h}\right) E\left[ f_{11,ij}^{(b)} \overline{f_{11,ij}^{(b)}} + \right. \\
&\quad \left. f_{22,ii}^{(b)} \overline{f_{11,jj}^{(b)}} - f_{12,ii}^{(b)} \overline{f_{12,jj}^{(b)}} - f_{21,ii}^{(b)} \overline{f_{21,jj}^{(b)}} \right] d\omega + o(1)
\end{aligned}$$

where for  $m_1, m_2 = 1, 2$ ,  $f_{m_1 m_2, ij}^{(b)}$  denotes the  $(i, j)^{th}$  entry of the  $p \times p$  matrix  $f_{m_1 m_2}^{(b)}$ .

The expression on the right hand side is asymptotically equivalent to

$$\frac{1}{h^{1/2}} \int_{-\pi}^{\pi} K^2(\lambda) d\lambda \int_{-\pi}^{\pi} \sum_{i_1, i_2=1}^2 (-1 + 2\delta_{i_1 i_2}) |tr(f_{i_1 i_2}^{(b)}(\omega))|^2 d\omega = \frac{\mu_0}{h^{1/2}}$$

For the variance we have

$$\begin{aligned}
Var\left(2\pi N h^{1/2} \hat{D}(b)\right) &= 4\pi^2 N^2 h Var\left(\hat{D}(b)\right) = \\
\frac{4\pi^2 h}{N^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{i,j,k,l=1}^p \sum_{k_1, k_2, k_3, k_4} K_h(\omega - \omega_{k_1}) K_h(\omega - \omega_{k_2}) K_h(\lambda - \omega_{k_3}) K_h(\lambda - \omega_{k_4}) \times \\
&\quad \left\{ E\left[ \left( I_{11,ij}^{(b)}(\omega_{k_1}) - I_{22,ij}^{(b)}(\omega_{k_1}) \right) \overline{\left( I_{11,ij}^{(b)}(\omega_{k_2}) - I_{22,ij}^{(b)}(\omega_{k_2}) \right)} \times \right. \right. \\
&\quad \left. \left. \overline{\left( I_{11,kl}^{(b)}(\omega_{k_3}) - I_{22,kl}^{(b)}(\omega_{k_3}) \right)} \left( I_{11,kl}^{(b)}(\omega_{k_4}) - I_{22,kl}^{(b)}(\omega_{k_4}) \right) \right] - \\
&\quad E\left[ \left( I_{11,ij}^{(b)}(\omega_{k_1}) - I_{22,ij}^{(b)}(\omega_{k_1}) \right) \overline{\left( I_{11,ij}^{(b)}(\omega_{k_2}) - I_{22,ij}^{(b)}(\omega_{k_2}) \right)} \right] \times \\
&\quad \left. E\left[ \left( I_{11,kl}^{(b)}(\omega_{k_3}) - I_{22,kl}^{(b)}(\omega_{k_3}) \right) \overline{\left( I_{11,kl}^{(b)}(\omega_{k_4}) - I_{22,kl}^{(b)}(\omega_{k_4}) \right)} \right] \right\} d\omega d\lambda
\end{aligned}$$

For a given  $i, j, k, l$ , the difference in expectations will result in an asymptotically non-zero term only when  $\omega_{k_1} = \omega_{k_3} \neq \omega_{k_2} = \omega_{k_4}$ ,  $\omega_{k_1} = -\omega_{k_3} \neq \omega_{k_2} = -\omega_{k_4}$ ,  $\omega_{k_1} = \omega_{k_4} \neq \omega_{k_2} = \omega_{k_3}$ ,  $\omega_{k_1} = -\omega_{k_4} \neq \omega_{k_2} = -\omega_{k_3}$ . In all the 4 cases, the product of expectation terms vanishes and each case results in the same contribution. Considering only the first case with a factor of

4 we have

$$\begin{aligned}
\text{Var}\left(2\pi N h^{1/2} \hat{D}(b)\right) &= \frac{16\pi^2 h}{N^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{k_1, k_2} \sum_{i, j, k, l=1}^p \frac{1}{h^4} K\left(\frac{\omega - \omega_{k_1}}{h}\right) K\left(\frac{\lambda - \omega_{k_1}}{h}\right) \times \\
&K\left(\frac{\omega - \omega_{k_2}}{h}\right) K\left(\frac{\lambda - \omega_{k_2}}{h}\right) \left[ f_{11,il}^{(b)}(\omega_{k_1}) \overline{f_{11,jk}^{(b)}(\omega_{k_1})} + f_{22,il}^{(b)}(\omega_{k_1}) \overline{f_{22,jk}^{(b)}(\omega_{k_1})} - \right. \\
&f_{12,il}^{(b)}(\omega_{k_1}) \overline{f_{12,jk}^{(b)}(\omega_{k_1})} - f_{21,il}^{(b)}(\omega_{k_1}) \overline{f_{21,jk}^{(b)}(\omega_{k_1})} \left. \right] \times \left[ \overline{f_{11,il}^{(b)}(\omega_{k_2})} f_{11,jk}^{(b)}(\omega_{k_2}) \right. \\
&\overline{f_{22,il}^{(b)}(\omega_{k_2})} f_{22,jk}^{(b)}(\omega_{k_2}) - \overline{f_{12,il}^{(b)}(\omega_{k_2})} f_{12,jk}^{(b)}(\omega_{k_2}) - \\
&\left. \overline{f_{21,il}^{(b)}(\omega_{k_2})} f_{21,jk}^{(b)}(\omega_{k_2}) \right] d\omega d\lambda + o(1) \\
&= \frac{16\pi^2 h}{N^2 h^4} \sum_{k_1, k_2} \left( \int_{-\pi}^{\pi} K\left(\frac{\omega - \omega_{k_1}}{h}\right) K\left(\frac{\omega - \omega_{k_2}}{h}\right) \right)^2 \sum_{i, j, k, l=1}^p \left[ \overline{f_{11,il}^{(b)}(\omega_{k_1})} f_{11,jk}^{(b)}(\omega_{k_1}) + \right. \\
&f_{22,il}^{(b)}(\omega_{k_1}) \overline{f_{22,jk}^{(b)}(\omega_{k_1})} - f_{12,il}^{(b)}(\omega_{k_1}) \overline{f_{12,jk}^{(b)}(\omega_{k_1})} - f_{21,il}^{(b)}(\omega_{k_1}) \overline{f_{21,jk}^{(b)}(\omega_{k_1})} \left. \right] \times \\
&\left[ \overline{f_{11,il}^{(b)}(\omega_{k_2})} f_{11,jk}^{(b)}(\omega_{k_2}) + \overline{f_{22,il}^{(b)}(\omega_{k_2})} f_{22,jk}^{(b)}(\omega_{k_2}) - \right. \\
&\left. \overline{f_{12,il}^{(b)}(\omega_{k_2})} f_{12,jk}^{(b)}(\omega_{k_2}) - \overline{f_{21,il}^{(b)}(\omega_{k_2})} f_{21,jk}^{(b)}(\omega_{k_2}) \right] d\omega d\lambda + o(1)
\end{aligned}$$

and this is asymptotically equivalent to

$$\begin{aligned}
&4 \int_{-2\pi}^{2\pi} \left( \int_{-\pi}^{\pi} K(\nu) K(\nu + z) d\nu \right)^2 dz \sum_{i_1, i_2, i_3, i_4=1}^2 (-1 + 2\delta_{i_1 i_2}) (-1 + 2\delta_{i_3 i_4}) \times \\
&|tr(f_{i_1 i_3}^{(b)}(\omega) \overline{f_{i_2 i_4}^{(b)}(\omega)}^T)|^2 d\omega = \sigma_0^2
\end{aligned}$$

□

**Proof of Theorem 2.7.2.** With the standardized entity  $Q_N$  and Assumption 2 on the bandwidth

$h$  we have, following Theorem 5.1 of Eichler (2008),

$$\begin{aligned} \frac{Q_N}{Nh^{1/2}} &\xrightarrow{P} A_1 \text{ for some } A_1 \in \mathbb{R}^+ \\ \Rightarrow \frac{2\pi N\hat{D}(b) - \hat{\mu}/h}{\hat{\sigma}/\sqrt{h}} &\xrightarrow{P} A_1 \\ \Rightarrow 2\pi\hat{D}(b) &\xrightarrow{P} A_2 \text{ where } A_2 \in \mathbb{R}^+ \end{aligned}$$

□

**Proof of Theorem 3.2.1.** Let  $\{X_t\}$  be the observed  $p$ -variate process generated by  $d$  stationary and  $p - d$  nonstationary sources as in (3.1). We have  $\mathcal{B} \subset \mathbb{R}^{d \times p}$  as the space of solutions for optimizing (3.7). Recall from (3.11) the constraints defining  $\mathcal{B}$  are given by

$$\begin{aligned} b_j V_r b_j^\top &= \delta(T), \\ B_1 B_1^\top &= I_d \end{aligned} \tag{A.1}$$

where  $b_j$ ,  $j = 1, 2, \dots, d$ , denotes the row vectors of any  $B_1 \in \mathcal{B}$  and  $r = 1, 2, \dots, m$  and  $\delta(T) \rightarrow 0$  as  $T \rightarrow \infty$ . Now, in order to uniquely identify  $\mathcal{C}(A_s)$ , we need for any  $W_1, W_2 \in \mathcal{B}$ ,

$$\mathcal{C}(W_1^\top) = \mathcal{C}(W_2^\top). \tag{A.2}$$

Following the proof of identifiability in von Büнау et al. (2009a), von Büнау et al. (2009b) and Theorem 4 in Hara et al. (2012), (A.2) holds if any  $p$ -vector  $w \in \mathcal{C}(W_1^\top)$  has degrees of freedom less than  $d$ . The  $m$  quadratic constraints in (A.1) account for reduction of  $m$  degrees of freedom. Hence the degrees of freedom of  $w = p - m < d$ . This implies  $m > p - d$ .

□

**Proof of Theorem 3.2.2.** Let  $d_0$  be the true dimension of the stationary subspace and we have

from (3.1)

$$X_t = A_s Y_t^s + A_n Y_t^n$$

where  $A_s$  is  $d_0 \times p$  and  $A_n$  is  $(p - d_0) \times p$ . Let  $\hat{d}$  be the estimated dimension given by (3.16) and  $\hat{B}_1 \in \mathbb{R}^{\hat{d} \times p}$  be the estimated transformation matrix. Here  $\hat{B}_1 \in \mathcal{B}$ , the space of feasible solutions defined by the constraints in (3.11) and the orthonormality constraint  $\hat{B}_1 \hat{B}_1^\top = I_{\hat{d}}$ . Now with  $\hat{B}_1$ , the estimate of the stationary source can be expressed as,

$$\hat{B}_1 X_t = \hat{B}_1 A_s Y_t^s + \hat{B}_1 A_n Y_t^n. \quad (\text{A.3})$$

To determine the probability  $P(\hat{d} = d_0)$  as  $T \rightarrow \infty$ , we look at various cases when  $\hat{d} \neq d_0$ , and determine their probabilities as  $T \rightarrow \infty$ . More precisely, we look at the probabilities  $P(\hat{d} = \tilde{d})$  when  $\tilde{d} = 1, 2, \dots, d_0 - 1, d_0 + 1, \dots, p$  as  $T \rightarrow \infty$ . Also, recall that for  $\tilde{d} = 1, 2, \dots, p$ ,  $(\mathcal{T}(m, n, \tilde{d}), c(\alpha_{\tilde{d}}))$  denotes the test statistic and critical value at level  $\alpha_{\tilde{d}}$  for the test of stationarity from Section 3.2.4. The critical value,  $c(\alpha_{\tilde{d}})$ , for a given  $\tilde{d}$  is the  $(1 - \alpha_{\tilde{d}})^{th}$  percentile of the  $\chi_{mn\tilde{d}(\tilde{d}+1)}^2$  distribution.

(i). First, we consider the simplest case with  $d_0 = 0$ . Here for any  $\tilde{d} > 0$  we have,

$$\begin{aligned} P(\hat{d} = \tilde{d}) &= P\left(\bigcap_{i=1}^{\tilde{d}} \{\mathcal{T}(m, n, i) < c(\alpha_i)\} \cap \{\mathcal{T}(m, n, \tilde{d} + 1) > c(\alpha_{\tilde{d}+1})\}\right) \\ &\leq P(\mathcal{T}(m, n, 1) < c(\alpha_1)) \end{aligned} \quad (\text{A.4})$$

Now if  $d_0 = 0$ ,  $\hat{B}_1 X_t \in \mathbb{R}^{\tilde{d} \times p}$ , where  $r > 0$ , is nonstationary and under Assumption 1 that  $\{X_t\}$  is a locally stationary Gaussian time series we have  $\mathcal{T}(m, n, 1) = O_p(T)$ ; see (Jentsch and

Subba Rao, 2015a, Section 5). Hence, when  $m, n$  are bounded positive integers, we have

$$\begin{aligned}
\lim_{T \rightarrow \infty} P\left(\mathcal{T}(m, n, 1) < c(\alpha_1)\right) &= 0 \\
\implies \lim_{T \rightarrow \infty} P\left(\hat{d} > d_0\right) &= 0 \\
\implies \lim_{T \rightarrow \infty} P\left(\hat{d} = d_0 = 0\right) &= 1
\end{aligned} \tag{A.5}$$

and this establishes part (a).

(ii). Next we take the case  $d_0 \in \{1, 2, \dots, p-1\}$ . Here for any  $\tilde{d} > d_0$  we have,

$$\begin{aligned}
P\left(\hat{d} = \tilde{d}\right) &= P\left(\bigcap_{i=1}^{d_0} \{\mathcal{T}(m, n, i) < c(\alpha_i)\} \times \bigcap_{j=d_0+1}^{\tilde{d}} \{\mathcal{T}(m, n, j) < c(\alpha_j)\} \right. \\
&\quad \left. \bigcap \{\mathcal{T}(m, n, \tilde{d}+1) > c(\alpha_{\tilde{d}+1})\} \right) \leq P\left(\mathcal{T}(m, n, d_0+1) < c(\alpha_{d_0+1})\right).
\end{aligned} \tag{A.6}$$

When  $\tilde{d} > d_0$ , it can be seen that  $\hat{B}_1 A_n \neq 0$  as the row vectors in  $\hat{B}_1$  include some orthonormal vectors from the space orthogonal to  $\mathcal{C}(A_s)$ . Hence  $\hat{B}_1 X_t$  is nonstationary and under Assumption 1 we have  $\mathcal{T}(m, n, d_0+1) = O_p(T)$ . Therefore for  $\tilde{d} > d_0$ ,

$$\begin{aligned}
\lim_{T \rightarrow \infty} P\left(\mathcal{T}(m, n, d_0+1) < c(\alpha_{d_0+1})\right) &= 0 \\
\implies \lim_{T \rightarrow \infty} P\left(\hat{d} > d_0\right) &= 0 \\
\implies \lim_{T \rightarrow \infty} P\left(\hat{d} \leq d_0\right) &= 1
\end{aligned} \tag{A.7}$$

and this establishes part (c).

(iii). Next we take the case  $d_0 \in \{1, 2, \dots, p\}$ , and case (b) applies. Now for  $\tilde{d} < d_0$  we have,

$$\begin{aligned}
P\left(\hat{d} = \tilde{d}\right) &= P\left(\bigcap_{i=1}^{\tilde{d}} \{\mathcal{T}(m, n, i) < c(\alpha_i)\} \bigcap \{\mathcal{T}(m, n, \tilde{d}+1) > c(\alpha_{\tilde{d}+1})\} \right. \\
&\quad \left. \leq P\left(\mathcal{T}(m, n, \tilde{d}+1) > c(\alpha_{\tilde{d}+1})\right)\right).
\end{aligned} \tag{A.8}$$

Now if  $\tilde{d} < d_0$ ,  $\hat{B}_1 X_t \in \mathbb{R}^{\tilde{d} \times p}$ , is stationary and  $\mathcal{T}(m, n, \tilde{d}) \xrightarrow{D} \chi_{mn\tilde{d}(\tilde{d}+1)}^2$ . Therefore for  $\tilde{d} < d_0$ ,

$$\begin{aligned}
\lim_{T \rightarrow \infty} P(\hat{d} = \tilde{d}) &\leq \lim_{T \rightarrow \infty} P(\mathcal{T}(m, n, \tilde{d} + 1) > c(\alpha_{\tilde{d}+1})) = \alpha_{\tilde{d}+1} \\
\implies \lim_{T \rightarrow \infty} P(\hat{d} = d_0) &= 1 - \lim_{T \rightarrow \infty} \sum_{j=0}^{d_0-1} P(\hat{d} = j) \\
&\geq 1 - \sum_{j=1}^{d_0} \alpha_j,
\end{aligned} \tag{A.9}$$

and this establishes part (b).

□

## APPENDIX B

### B.1 Optimization Details from Section 3.2.1

Here we provide more details on the optimization problem formulated in Section 3.2. We begin by deriving the expression of the gradient matrix. We denote  $I^d$  as a  $d \times p$  matrix with rows from the first  $d$  rows of a  $p \times p$  identity matrix  $I_p$ .

Recall that the lag- $r$  DFT covariance from (3.6) of the series  $Y_t^s = B_1 X_t = I^d B X_t$  can be written as

$$\begin{aligned}
 \hat{\Gamma}_r^{Y^s} &= \frac{1}{T} \sum_{k=1}^T J(\omega_k) J(\omega_{k+r})^* = \frac{1}{T} \sum_{k=1}^T \left( \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T Y_t^s \exp(-it\omega_k) \right) \times \\
 &\quad \left( \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T Y_t^s \exp(-it\omega_{k+r}) \right)^* \\
 &= \frac{1}{2\pi T^3} \sum_{k=1}^T \left( \sum_{t=1}^T I^d B X_t \exp(-it\omega_k) \right) \times \left( \sum_{t=1}^T I^d B X_t \exp(it\omega_{k+r}) \right)^* \\
 &= \frac{1}{2\pi T^3} \sum_{k=1}^T \sum_{t_1, t_2=1}^T I^d B X_{t_1} X_{t_2}^\top B^\top (I^d)^\top \exp(-it_1\omega_k) \exp(it_2\omega_{k+r}) \\
 &= B_1 \left( \frac{1}{2\pi T^3} \sum_{k=1}^T \sum_{t_1, t_2=1}^T X_{t_1} X_{t_2}^\top \exp(-it_1\omega_k) \exp(it_2\omega_{k+r}) \right) B_1^\top \\
 &= B_1 (U_r) B_1^\top = B_1 (U_r^R + iU_r^I) B_1^\top
 \end{aligned}$$

where  $B_1 = I^d B$ . Let  $U_r^R, U_r^I$  are the real and imaginary parts of  $U_r$  respectively. Our objective

function defined in (3.7) can be written as

$$\begin{aligned}
D_Y(B) &= \sum_{r=1}^m \|B_1 U_r^R B_1^\top\|_F^2 + \|B_1 U_r^I B_1^\top\|_F^2 \\
&= \sum_{r=1}^m \text{tr}(B_1 (U_r^R)^\top B_1^\top B_1 U_r^R B_1^\top) + \text{tr}(B_1 (U_r^I)^\top B_1^\top B_1 (U_r^I) B_1^\top) \\
&= \sum_{r=1}^m \text{tr}(I^d B (U_r^R)^\top B^\top (I^d)^\top I^d B U_r^R B^\top (I^d)^\top) + \text{tr}(I^d B (U_r^I)^\top B^\top (I^d)^\top I^d B U_r^I B^\top (I^d)^\top)
\end{aligned}$$

where  $\text{tr}(\cdot)$  denotes trace of a matrix. In order to take the derivative of the above expression with respect to  $B$  we make use of some results on derivatives of the trace in Petersen and Pedersen (2012). First, for a given  $r$ , we consider the first term  $\text{tr}(I^d B (U_r^R)^\top B^\top (I^d)^\top I^d B U_r^R B^\top (I^d)^\top)$  and denote it as  $D_{1,r}(B)$  and the corresponding  $p \times p$  gradient matrix as  $G_{1,r}$ . We have

$$\begin{aligned}
G_{1,r} &= (I^d)^\top I^d B (U_r^R)^\top B^\top (I^d)^\top I^d B U_r^R + (I^d)^\top I^d B U_r^R B^\top (I^d)^\top I^d B (U_r^R)^\top \\
&\quad + (I^d)^\top I^d B U_r^R B^\top (I^d)^\top I^d B (U_r^R)^\top + (I^d)^\top I^d B (U_r^R)^\top B^\top (I^d)^\top I^d B U_r^R
\end{aligned}$$

Similarly, the second term  $D_{2,r}(B) = \text{tr}(I^d B (U_r^I)^\top B^\top (I^d)^\top I^d B U_r^I B^\top (I^d)^\top)$  has a gradient  $G_{2,r}$ . Hence the gradient for  $D_Y(B)$  is given by  $\frac{\partial D_Y(B)}{\partial B} = G = \sum_{r=1}^m (G_{1,r} + G_{2,r})$ . Now, with the space of orthogonal matrices being parameterized as matrix exponentials of skew-symmetric matrices, we have  $B = e^{c \times H}$  where  $H$  is a skew-symmetric matrix and  $c \in \mathbb{R}$ . The gradient descent algorithm to obtain a solution for the optimization problem in (3.7) involves the following steps:

1. Initialize  $B_0$ , a random  $p \times p$  orthogonal matrix with positive determinant using the technique from Stewart (1980).
2. Find the next search direction as  $H = G B_0^\top - B_0 G^\top$ .
3. Optimize  $D_Y(e^{c \times H} B_0)$  w.r.t  $c$  where  $c \in \mathbb{R}$ .



4. Obtain  $B_{(n)}$  as  $B_{(n)} = e^{c \times H} B_0$  and check for convergence by examining values of successive iterates.
5. Set  $B_0 = B_{(n)}$  and repeat Steps 2,3,4 until convergence.

The presence of the constant  $c$  above ensures that during iterations, larger steps are taken in the Lie algebra of skew-symmetric matrices thereby speeding up the iterative process; see Section 8.1 of Plumbley (2005).

Note that in Step 1 we consider 10 different initial guesses of random orthogonal matrices and determine the final solution that yields the least objective function value.

## B.2 Details from Section 3.2.3

Recall that  $U_r^R$  and  $U_r^I$  are the real and imaginary parts of the  $p \times p$  matrix  $\hat{\Gamma}_r^X$ . We now write for  $j = 1, 2, \dots, d$ ,

$$\begin{aligned} |Re(b_j \hat{\Gamma}_r^X b_j^T)| + |Im(b_j \hat{\Gamma}_r^X b_j^T)| &= |b_j U_r^R b_j^T| + |b_j U_r^I b_j^T| \\ &= |b_j (P_r^R D_r^R (P_r^R)^T) b_j^T| + |b_j (P_r^I D_r^I (P_r^I)^T) b_j^T| \end{aligned}$$

where  $|\cdot|$  denotes the absolute value and,

$$U_r^R = P_r^R D_r^R (P_r^R)^T \quad \text{and} \quad U_r^I = P_r^I D_r^I (P_r^I)^T \quad (\text{B.1})$$

denotes the eigen decomposition of the  $p \times p$  symmetric matrices  $U_r^R$  and  $U_r^I$  respectively.

We write this decomposition as a difference of non-negative definite matrices i.e

$$\begin{aligned} P_r^R D_r^R (P_r^R)^T &= P_r^R D_r^{R+} (P_r^R)^T - P_r^R D_r^{R-} (P_r^R)^T \\ P_r^I D_r^I (P_r^I)^T &= P_r^I D_r^{I+} (P_r^I)^T - P_r^I D_r^{I-} (P_r^I)^T \end{aligned}$$

where for  $i = 1, 2, \dots, p$ , the  $p \times p$  diagonal matrices  $D_r^{R+}$ ,  $D_r^{R-}$  and  $D_r^{I+}$ ,  $D_r^{I-}$  are defined as

$$\begin{aligned} D_{r,ii}^{R+} &= \begin{cases} D_{r,ii}^R, & \text{if } D_{r,ii}^R > 0. \\ 0, & \text{otherwise.} \end{cases}, & D_{r,ii}^{I+} &= \begin{cases} D_{r,ii}^I, & \text{if } D_{r,ii}^I > 0. \\ 0, & \text{otherwise.} \end{cases} \\ D_{r,ii}^{R-} &= \begin{cases} -D_{r,ii}^R, & \text{if } D_{r,ii}^R < 0. \\ 0, & \text{otherwise.} \end{cases}, & D_{r,ii}^{I-} &= \begin{cases} -D_{r,ii}^I, & \text{if } D_{r,ii}^I < 0. \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{B.2})$$

Hence we have, for  $r = 1, 2, \dots, m$  and  $j = 1, 2, \dots, d$ ,

$$\begin{aligned} |\operatorname{Re}(b_j \hat{\Gamma}_r^X b_j^\top)| + |\operatorname{Im}(b_j \hat{\Gamma}_r^X b_j^\top)| &= |b_j (P_r^R D_r^{R+} (P_r^R)^\top) b_j^\top - b_j (P_r^R D_r^{R-} (P_r^R)^\top) b_j^\top| + \\ &\quad |b_j (P_r^I D_r^{I+} (P_r^I)^\top) b_j^\top - b_j (P_r^I D_r^{I-} (P_r^I)^\top) b_j^\top| \\ &\leq |b_j (P_r^R D_r^{R+} (P_r^R)^\top) b_j^\top| + |b_j (P_r^R D_r^{R-} (P_r^R)^\top) b_j^\top| + \\ &\quad |b_j (P_r^I D_r^{I+} (P_r^I)^\top) b_j^\top| + |b_j (P_r^I D_r^{I-} (P_r^I)^\top) b_j^\top| \\ &= b_j V_r b_j^\top. \end{aligned} \quad (\text{B.3})$$

where  $V_r = (P_r^R D_r^{R+} (P_r^R)^\top + P_r^R D_r^{R-} (P_r^R)^\top) + (P_r^I D_r^{I+} (P_r^I)^\top + P_r^I D_r^{I-} (P_r^I)^\top)$ .