DETECTING CHANGE-POINTS IN THE MEAN OF MULTIVARIATE TIME SERIES

DETECTING CHANGE-POINTS IN THE MEAN OF MULTIVARIATE TIME SERIES

By RAMKRISHNA JYOTI SAMANTA, B.Math

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2024)

Hamilton, Ontario, Canada (Mathematics and Statistics)

TITLE:	Detecting change-points in the mean of multivariate time series
AUTHOR:	Ramkrishna Jyoti Samanta B.Math (Mathematics & Statistics), Indian Statistical Institute.
SUPERVISOR:	Professor Narayanaswamy Balakrishnan
NUMBER OF PAGES:	x, 63

Abstract

After providing a detailed literature review of the change-point detection methods, this work delves into presenting a probabilistic method for analyzing linear process data with dependent innovations, focusing on detecting change-points in the mean and estimating its spectral density. We develop a test for identifying change-points in the mean of the data, aiming to detect shifts in the underlying distribution. Additionally, we propose a consistent estimator for the spectral density of the data, contingent upon fundamental assumptions, notably the long-run variance. By leveraging probabilistic techniques, our approach provides reliable tools for understanding temporal changes in linear process data. Through theoretical analysis and empirical evaluation, we demonstrate the efficacy and consistency of our proposed methods, offering valuable insights for practitioners in various fields dealing with time series data analysis.

Acknowledgements

I am immensely grateful to my advisor, Professor N. Balakrishnan, for his unwavering support, invaluable guidance and granting me the freedom to pursue my own ideas. Under his mentorship, I have been fortunate to engage in several projects even before joining McMaster University, which I believe, have provided me with invaluable research experience and will continue to prove beneficial to me in my research career. His mentorship, immense knowledge and mathematical skills have always motivated me, and I owe a great deal of my achievements to his guidance.

Additionally, I extend my heartfelt thanks to Professor Shui Feng for his constant inspiration and motivational encouragement. His insights have been a source of inspiration, pushing me to strive for excellence in my work.

I would also like to express my gratitude to Professor Noah Forman for his guidance and assistance in navigating the complexities of the book on *Convergence of Probability Measures*. His expertise and patience have been instrumental in helping me comprehend and complete this challenging work.

And finally I would like to thank Professor Katherine Davies for her fine reading of the thesis and her valuable suggestions in improving the readability of the thesis to a great extent.

Without the support and mentorship of Professors Balakrishnan, Feng, Forman and Davies, I would not have been able to achieve what I have. Their guidance and encouragement have been invaluable assets throughout my academic journey.

Table of Contents

List of Figures

List of Tables

Notation

$\|x\|_p$	The L_p norm of the vector x
$\|A\|$	The matrix norm of the matrix A
$P(\cdot)$	The probability measure function
$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$\mathbb{I}(\cdot)$	The indicator function
$\lfloor x \rfloor$	The greatest integer less than or equal to x

Declaration of Academic Achievement

I declare that the contents of this thesis are my own.

Chapter 1

Introduction

A structural break refers to a sudden and significant change in the underlying data-generating process of a time series. This change can manifest in various parameters of the model, such as the mean, variance, or autoregressive coefficients and a change-point refers to a specific time point within a time series where a structural break occurs. For example, consider a temperature dataset spanning several decades. Suppose the data exhibits a stable pattern with gradual temperature increases over time, representing a consistent warming trend. However, in recent years, there has been a sudden and persistent drop in temperatures, leading to colder conditions. The time point where this abrupt change occurs marks the change-point, indicating a structural break in the temperature trend. Detecting structural breaks has captivated researchers' interest since Pagetest [13] proposed a test for it as early as in 1955. Identifying changes in the mean of a multivariate data is a fundamental problem of interest in various disciplines such as signal processing, finance, environmental monitoring, and healthcare. Change-point detection methods are pivotal in pinpointing abrupt shifts or structural breaks in the underlying distribution of multivariate data over time or space. These methods are indispensable for comprehending and analyzing dynamic systems, detecting anomalies, and making well-informed decisions based on evolving data patterns. Various change-point detection methods have been discussed briefly in Kolz [14], where an overview is presented of various techniques used by researchers in developing methods for change-point detection.

Multivariate change-point detection presents unique challenges compared to the univariate scenario due to the interdependence among the observed variables. In the case of multivariate data, changes in one variable may correlate with changes in others, necessitating sophisticated techniques to accurately detect shifts in the mean across multiple dimensions. A time series $\{X_t\}$ is called a linear process if it can be expressed as:

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},$$

, where μ is the mean of X_t, $\{\psi_j\}$ are the coefficients with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, and $\{\varepsilon_t\}$ is white noise with zero mean and constant variance. In certain types of data, it will be feasible to model data with some level of dependence among its coordinates. This allows for the introduction of dependence in the innovations. In our case, we specifically consider linear process data with m-dependent $\{\varepsilon_t\}$'s. A sequence of random variables $\{X_t\}_{t\in\mathbb{Z}}$ is said to be m-dependent if for any two sets of indices I and J such that the minimum distance between elements of I and elements of J is greater than m, the sets of random variables $\{X_i : i \in I\}$ and $\{X_j : j \in J\}$ are independent. Formally,

$$\{X_i : i \in I\} \perp\!\!\!\perp \{X_j : j \in J\} \quad \text{whenever} \quad \min_{i\in I, j\in J} |i - j| > m$$

Here, $\perp\!\!\!\perp$ denotes the independence of the two sets of random variables. To conduct a change-point analysis, we first need to perform hypothesis testing to determine whether a change-point exists at all. This typically involves a statistic dependent on N, the length of the data, whose distribution for finite N is unknown. Hence, we consider data with a large length ($N \to \infty$), yielding a suitable statistic with known critical values from Kiefer [8]. Moreover, Zeileis [18] stated that if a time series is observed for a long enough period of time, there would be some economic or political or climatic factors that would cause the structure of the series to change at some points. This also entails estimating the long-run covariance matrix of the data, which leads to the estimation of the spectral density of the data.

Let $\{X_t\}$ be a stationary time series with mean μ. The spectral density function $f(\lambda)$ at frequency λ is defined as the Fourier transform of the autocovariance function $\Gamma(h)$. For a discrete-time stationary process, the spectral density is given by:

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-ih\lambda}, \quad -\pi \leq \lambda \leq \pi.$$

Estimating spectral density has been studied extensively for univariate linear process data with independent innovations, as can be seen in Brockwell [3]. Here, we aim to derive a consistent estimator of the spectral density for such data, especially to estimate the long-run covariance matrix of the data. This, in turn, facilitates us to provide an estimator for the change-point under the assumption that there is a change-point. As usual, we consider a d dimensional time series $X_t = (X_t^1, \ldots, X_t^d)$, and $t \in \{1, \ldots, N\}$ corresponding to the time-points of observations. In our case, we have:

$$H_0: \quad \mathbb{E}[X_1] = \ldots = \mathbb{E}[X_N]$$

vs

$$H_a: \quad \exists \; T^* \text{ such that } \mathbb{E}[X_1] = \ldots = \mathbb{E}[X_{T^*}] \neq \mathbb{E}[X_{T^*+1}] \ldots = \mathbb{E}[X_N].$$

This is the scenario when we have a single change-point in our data. There have been many works in the literature discussing various types of change-point detection methods in multivariate data. Some of them include Amiri [1], Viviani [17], Meier [12], and Zeileis [18], etc. In the next chapter, we present a short review of some of the existing methods in this direction.

Chapter 2

Preliminaries

This chapter presents some preliminary details that are required mainly in Chapter 3 and 4, where we discuss some background literature and theoretical results, respectively.

2.1 Some notions on modes of convergence

1. **Almost Sure Convergence (a.s.):** Let (X_n) and X be (S, ρ)-valued random variables defined on the same probability space. Then, we say that X_n converges almost surely to X, denoted by $X_n \xrightarrow{\text{a.s.}} X$, if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

or, if for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\sup_{m>n} \rho(X_n, X) < \epsilon) = 1.$$

2. **Convergence in Probability:** Let (X_n) and X be (S, ρ)-valued random variables. Then, we say that X_n converges to X in probability, denoted by $X_n \xrightarrow{P} X$, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(\rho(X_n, X) > \epsilon) = 0.$$

3. **Convergence in p^{th} Moment:** Let (X_n) and X be (S, ρ)-valued random variables. Then, we say that X_n converges to X in p^{th} moment, denoted by $X_n \xrightarrow{L^p} X$, if

$$\lim_{n \to \infty} \mathbb{E}[\rho(X_n, X)^p] = 0$$

for some $p > 0$.

4. **Weak Convergence of Random Variables:** Let (X_n) and X be (S, ρ)-valued random variables. Then, we say that X_n weakly converges to X, denoted by $X_n \xrightarrow{w} X$, if for all bounded and continuous functions $f : \mathbb{R} \to \mathbb{R}$, we have

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} f(x) dP_{X_n}(x) = \int_{-\infty}^{\infty} f(x) dP_X(x),$$

where $P_Y(.)$ is the probability measure corresponding to the random variable Y.

It is well-known that the following implications hold among the above concepts:

- Convergence in Probability $\Rightarrow$ Weak convergence.
- Almost Sure Convergence $\Rightarrow$ Convergence in Probability
- Convergence in p^{th} Moment $\Rightarrow$ Convergence in Probability

Definition 2.1. [Mahalanobis distance] Given a vector $\mathbf{x}$ and a multivariate distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the Mahalanobis distance D_M of $\mathbf{x}$ from the mean $\boldsymbol{\mu}$ is defined as:

$$D_M(\mathbf{x}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}.$$

Lemma 2.2. *[Cramér-Wold Device] Let $X_1, X_2, \ldots$ be a sequence of d-dimensional random vectors, each defined on the same probability space $(\Omega, \mathcal{F}, P)$. The Cramér-Wold theorem states that if for every $\mathbf{a} \in \mathbb{R}^d$, the sequence of scalar products $\mathbf{a}^T X_n$ converges in distribution to a scalar random variable Y, then the random vector X_n converges in distribution to a random vector X such that $\mathbf{a}^T X$ has the same distribution as Y for all $\mathbf{a} \in \mathbb{R}^d$.*

Lemma 2.3. *[Portmanteau's Theorem] Billingsley [2] Let (X_n) and X be (S, ρ)-valued random variables. Then, the following statements are equivalent:*

1. *X_n weakly converges to X;*

2. $$\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \le \mathbb{P}(X \in F), \text{ for all closed sets } F;$$

3. $$\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \ge \mathbb{P}(X \in G), \text{ for all open sets } G;$$

4. $$\lim_{n\to\infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B), \text{ for all Borel sets } B \text{ such that } \mathbb{P}(X \in \partial B) = 0,$$

where ∂B is the boundary of the set $B \subseteq S$.

Lemma 2.4. *[Continuous Mapping Theorem] Billingsley [2] Let (X_n) be a sequence of (S, ρ)-valued random variables converging weakly (in distribution) to a random variable X, and $g : S \to \mathbb{R}$ be a function, with measure of set of discontinuities zero. Then, the sequence $(g(X_n))$ converges weakly (in distribution) to $g(X)$.*

Lemma 2.5. *Suppose X_n, X and Y_n's are all (S, ρ) valued random variables. If (X_n) converges weakly (in distribution) to X and $\rho(X_n, Y_n) \xrightarrow{P} 0$, then (Y_n) converges weakly (in distribution) to Y.*

Definition 2.6. [Uniformly tight] Billingsley [2] Suppose X_n's are all (S, ρ) valued random variables. (X_n) is said to be **uniformly tight** if, for every $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq S$ such that $P(X_n \in K_\epsilon) > 1 - \epsilon$ for all n.

Lemma 2.7. *[Prokhorov's Theorem] Billingsley [2] Let (X_n) be a sequence of a (S, ρ)-valued random variables, where S is complete and separable space, defined on a probability space $(\Omega, \mathcal{F}, P)$. Then, (X_n) contains a subsequence that converges weakly if and only if (X_n) is uniformly tight.*

In particular, for $\mathbb{R}^d$-valued sequence of random variables (X_n)'s, (X_n) contains a subsequence that converges weakly if and only if $(X_n)_i$ (i-th coordinate of the sequence) is uniformly tight $\forall 1 \leq i \leq d$.

Lemma 2.8. *[Minkowski's Inequality] Let $p \geq 1$ and X and Y be any p-integrable random variables. Then, we have*

$$(\mathbb{E}[|X + Y|^p])^{1/p} \leq (\mathbb{E}[||X||^p])^{1/p} + (\mathbb{E}[||Y||^p])^{1/p}.$$

Definition 2.9. [Brownian Motion] Billingsley [2] A **Brownian motion** $W(t)$ on the interval $[0,1]$ is a stochastic process satisfying the following properties:

1. $W(0) = 0$ almost surely;
2. For any $0 \leq s < t \leq 1$, the increment $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$;
3. The increments of $W(t)$ are independent;
4. $W(t)$ has continuous sample paths.

Definition 2.10. [Brownian Bridge] Billingsley [2] A **Brownian bridge** $W^0(t)$ on the interval $[0,1]$ is a stochastic process defined as the conditional distribution of a Brownian motion $W(t)$ given that $W(0) = W(1) = 0$. In other words, it is a Brownian motion conditioned to start and end at 0. The Brownian bridge $W^0(t)$ can be represented as:

$$W^0(t) = W(t) - tW(1),$$

where $W(t)$ is the standard Brownian motion.

Definition 2.11. [Ergodic Sequence] Durrett [4] Let $(\Omega, \mathcal{F}, P)$ be a probability space, and ϕ be a map that preserves P, meaning that $P(\phi^{-1}(A)) = P(A)$ for all $A \in \mathcal{F}$. Define $X_n(\omega) = X(\phi^n(\omega))$, where X is a random variable.

A set $A \in \mathcal{F}$ is said to be invariant if $\phi^{-1}(A) = A$. Let $\mathcal{I}$ be the collection of invariant events.

A measure-preserving transformation on $(\Omega, \mathcal{F}, P)$ is said to be ergodic if $\mathcal{I}$ is trivial, i.e., for every $A \in \mathcal{I}$, $P(A) \in \{0, 1\}$.

Lemma 2.12. *[Ergodic Theorem] Durrett [4] Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, and* $\phi : \Omega \to \Omega$ *be a measure-preserving map. Suppose* $X : \Omega \to \mathbb{R}$ *is an integrable random variable. The ergodic theorem states that for almost every* $\omega \in \Omega$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X(\phi^n(\omega)) = \mathbb{E}[X|I](\omega).$$

Proposition 2.13. *Stout [15] Let* $\{X_i, i \geq 1\}$ *be the coordinate representation of a stationary sequence. There exists a measure-preserving transformation* ϕ *on* $(\mathbb{R}_\infty, C_\infty, P)$ *such that* $X_1(\omega) = X_1(\omega)$, $X_2(\omega) = X(\phi(\omega))$, $\ldots$, $X_n(\omega) = X(\phi^n(\omega)), \ldots$ *for all* $\omega \in \mathbb{R}_\infty$.

Proposition 2.14. *[The pointwise ergodic theorem for stationary sequences] Stout [15] Let* $\{X_i, i \geq 1\}$ *be stationary with* $\mathbb{E}|X_1| < \infty$. *Then* $\sum_{i=1}^{n} X_i/n \to E[X_1|I]$ *a.s. If in addition* $X_i, i \geq 1$ *is ergodic,* $\sum_{i=1}^{n} X_i/n \to E[X_1]$ *a.s.*

Proposition 2.15. *Stout [15] Let* $\{X_i, i \geq 1\}$ *be stationary ergodic and* $\phi : R_\infty \to R$ *be measurable. Let* $Y_i = \phi(X_i, X_{i+1}, \ldots), \forall i \geq 1$. *Then* $Y_i, i \geq 1$ *is stationary ergodic.*

Lemma 2.16. *Stout [15] Let* $\{X_i, i \geq 1\}$ *be independent identically distributed. Then* $\{X_i, i \geq 1\}$ *is stationary ergodic.*

2.2 Some common metric spaces

1. **Space of continuous functions C[0, T]:** This is the space of all of continuous functions on the interval $[0, T]$ with the Borel-sigma algebra generated by the metric

$$d(f, g) = \sup_{t\in[0,T]} |f(t) - g(t)|,$$

where $f, g \in C[0, T]$.

2. **Space of cadlag functions $\boldsymbol{D}$[0, T]:** This is the space of right-continuous, with left-hand limit functions; that is,

   (i) For $0 \leq t \leq T$, $f(t^+) = \lim_{s \downarrow t} f(s)$ exists and $f(t^+) = f(t)$;

   (ii) For $0 \leq t \leq T$, $f(t^-) = \lim_{s \uparrow t} f(s)$ exists.

   For $f, g \in D[0, T]$, the metric $d_0(f, g)$ is defined to be the infimum of those positive ϵ for which there exists λ, which is strictly increasing, continuous mapping from $[0, T]$ onto itself with $\lambda(0) = 0$ and $\lambda(T) = T$, such that

$$\sup_t |\lambda(t) - t| \leq \epsilon$$

   and

$$\sup_t |f(t) - g(\lambda(t))| \leq \epsilon.$$

These metrics can be extended to $C[0, \infty]$ and $D[0, \infty]$ by defining

$$\hat{d}(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} d_n(f, g),$$

where $d_n(f, g)$ is the metric when f and g are restricted to the interval $[0, n]$.

Chapter 3

Literature review

In this chapter, we present a short literature review for categorizing the different methods available in the literature based on whether the structural break is due to changes in parameters of the linear regression models or Vector Autoregression (VAR) models, owing to changes in the mean, covariance or distribution function. The latter is further classified based on the approach employed cumulative sums, distance measures and some combination of these approaches.

3.1 Change in parameters of VAR

A Vector Autoregressive (VAR) model is a multivariate time series model used to describe the joint dynamics of multiple time series variables. In a VAR model, each variable is regressed on its own lagged values and the lagged values of all other variables in the system. Mathematically, a VAR model of order p for k variables is represented as:

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \cdots + A_p X_{t-p} + \varepsilon_t$$

where:

- X_t is a k-dimensional vector of variables at time t.
- $A_1, A_2, \ldots, A_p$ are coefficient matrices capturing the lagged relationships between variables.
- ε_t is a k-dimensional vector of error terms.

Malo [10] proposed a distribution-free method which employs an energy distance formula such as $\mu_\alpha(A, B)$ in (3.1.2) and (3.1.3), based on the Euclidean norm to decipher whether a change-point exists or not in a linear regression model with one dependent variable and several covariates. For a multiple linear regression model with Y_t as the dependent variable, X_t as the regressor, ϵ_t as the residuals and β_i as the coefficient values for each phase between the change-points, the model is described as $Y_t = X_t' \beta_i + \epsilon_t$, where $t = \tau_{i-1} + 1, \ldots, \tau_i$, for $i = 1, 2, \ldots, k+1$. Here, k denotes the number of change-points and τ_i, for $1 \leqslant i \leqslant k$, represents the location of those points. The unknown regression coefficients are estimated under regularized conditions. If $E_1, \ldots, E_{k+1}$ are the residuals obtained from the above regression in the $k+1$ regimes and the corresponding distribution functions are $F_1, \ldots, F_{k+1}$, then the null hypothesis to be tested is $H_0 : F_1 = \ldots = F_{k+1}$, against the alternative that at least one of the F_i's is not equal to the others. The test statistic is defined as

$$F_\alpha(E_1, \ldots, E_{k+1}) = \frac{S_\alpha(E_1, \ldots, E_{k+1})/k}{W_\alpha(E_1, \ldots, E_{k+1})/(N-k-1)}, \tag{3.1.1}$$

where α is a parameter that can be chosen. The numerator of (3.1.1) is equal to

$$S_\alpha(E_1, \ldots, E_{k+1}) = T_\alpha(E_1, \ldots, E_{k+1}) - W_\alpha(E_1, \ldots, E_{k+1}),$$

where

$$T_\alpha(E_1, \ldots, E_{k+1}) = \frac{N}{2}\mu_\alpha(E, E) \tag{3.1.2}$$

and

$$W_\alpha(E_1, \ldots, E_{k+1}) = \sum_{j=1}^{k+1} \binom{n_j}{2} \mu_\alpha(E_j, E_j), \tag{3.1.3}$$

where $\mu_\alpha(A, B) = \frac{1}{n1n2}\sum_{i=1}^{n1}\sum_{j=1}^{n2} |a_i - b_j|^\alpha$ for a_i's $\in A$ and b_j's $\in B$. To perform this test, random permutations are done. This test can be performed with slight modifications for the case of known and unknown change-points. The optimization is done mainly through dynamic programming by minimizing energy distances between residuals in each regime. This algorithm can be performed in software R using the package *changedetection*. To access the performance of this method, the statistic

$$R = \sum_{i=1}^{min(k,\hat{k})} |\tau_i - \hat{\tau}_i| + r\left|k - \hat{k}\right|, \tag{3.1.4}$$

where, τ_i's and $\hat{\tau}_i$ are the actual and estimated change-points, respectively and similarly k and $\hat{k}$ are the actual and estimated number of change-points respectively. Further, r is a penalty term associated with the change-point location error. The smaller the value of R, the better the method.

Gao [5] used a two-stage procedure to estimate the number and location of change-points in a piecewise stationary multivariate time series with a VAR setup. In the first step, assuming k change-points and $k+1$ phases subsequently, the piecewise VAR model for a p dimensional X_t for lag m is given by

$$X_t = B^j X'_{t-1} + \epsilon_t, \tag{3.1.5}$$

where $j = 1, \ldots, k+1$ represents the different regimes between the change-points, B^j is the $p \times pm$ matrix of coefficients for each phase and $X'_{t-1} = [X_{t-1}, X_{t-2}, \ldots, X_{t-m}]^\top$ is a $pm \times 1$ vector. Change-points are estimated separately for each of the p components and are combined together thereafter. The change-points are estimated using

$$\phi_{i,q} = \begin{cases} B^{j+1}_{i.} - B^j_{i.}, & q = \tau^i_j, \\ 0, & otherwise, \end{cases} \tag{3.1.6}$$

where τ^i_j represents change-points, $i = 1, \ldots, p$, $q = 2, \ldots, N$, and $\phi_{i,q}$ is a $1 \times pm$ vector. When $\phi_{i,q}$ has at least one non-zero entry, then there is a change in the parameter B_i, which is estimated as

$$\hat{\phi}_i = \arg\min_{\phi_i} \left(||X^0_i - X'_i \phi||^2 + \lambda_N \sum_{q=m+1}^{N} ||\phi_{i,q}||^2 \right). \tag{3.1.7}$$

These change-point estimates obtained as above are based on non-zero values of $\phi_{i,q}$ and are denoted by $\mathcal{A}_i$. The values of change-points thus obtained are subjected to the second step of further selection and estimation. In the selection stage, for the true change-point, τ_j for $j = 1, \ldots, k$, the change-points selected in the previous step for each of the components are grouped based on their proximity to the original one using the LASSO penalty function. Then, using the backward elimination algorithm, the most insignificant change-point (the one for which the value of the argument in (3.1.7) is lesser than a chosen cut off) is eliminated. Using the LASSO (Least Absolute Shrinkage and Selection Operator) and OLS (Ordinary Least Squares), the coefficients are estimated for the selected predictors.

3.2 Methods to detect changes in the mean or distribution

Based on cumulative sums

Let p be the number of entities (e.g., individuals, firms, countries) and N be the number of time periods. A typical panel dataset can be expressed as:

$$\{X_{t,i} : i = 1, 2, \ldots, p;\ t = 1, 2, \ldots, N\}.$$

For a single change in mean in a panel data, Horvath [7] considered the independent data model for each of the p variables,

$$X_{t,i} = \mu + \delta\, \mathbb{I}(t > t_0) + e_{t,i}, \tag{3.2.1}$$

where $t = 1, \ldots, N$, $i = 1, \ldots, p$ and t_0 is the suspected change-point location. The innovations, $e_{t,i}$'s follow a linear process with independent residuals. Then, the maximum of the absolute value of

$$\begin{aligned} \bar{C}_{p,N}(x) &= \frac{1}{p^{1/2}} \sum_{i=1}^{p} \left(\frac{1}{\sigma_i^2} Z_{N,i}^2(x) - \frac{\lfloor Nx \rfloor (N - \lfloor Nx \rfloor)}{N^2} \right), \\ \text{where } Z_{N,i}(x) &= \frac{1}{N^{1/2}} \left(S_{N,i}(x) - \frac{\lfloor Nx \rfloor}{N} S_{N,i}(1) \right), \\ S_{N,i}(x) &= \sum_{t=1}^{\lfloor Nx \rfloor} X_{t,N} \end{aligned} \tag{3.2.2}$$

is calculated, where $0 \leqslant x < 1$. The change-point location is at

$$\hat{t}_0 = \underset{1 \leqslant t \leqslant N-1}{\arg\max} \left(\sum_{i=1}^{p} \frac{Z_{N,i}^2(t/N)}{t(N-t)} \right). \tag{3.2.3}$$

Based on distance measures

Matteson [11] proposed using Euclidean distance for calculating a multivariate distance measure and the bisection method to estimate the number and location of change-points. This non-parametric method of change-point detection is named E-Divisive. An empirical measure of energy distance based on Euclidean distance Szekely [16] is considered for independent and identically distributed samples from random variables $\mathbf{X}$ and $\mathbf{Y}$ (dimensions of $\mathbf{X}$ and $\mathbf{Y}$ need not be equal) as follows:

$$\hat{\epsilon}(X_a, Y_b; \alpha) = \frac{2}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} |X_i - Y_j|^{\alpha} - \binom{a}{2}^{-1} \sum_{1 \leqslant i < k \leqslant < a} |X_i - X_k|^{\alpha} - \binom{b}{2}^{-1} \sum_{1 \leqslant j < k \leqslant < b} |Y_j - Y_k|^{\alpha}. \quad (3.2.4)$$

The scaled version of (3.2.4) is given by

$$\hat{\mathcal{Q}}(X_a, Y_b; \alpha) = \frac{ab}{a+b} \hat{\epsilon}(X_a, Y_b; \alpha). \quad (3.2.5)$$

To estimate one breakpoint $\hat{\tau}$, arg max of $\hat{\mathcal{Q}}(X_\tau, Y_\tau(k); \alpha)$ over τ is calculated. For several unknown change-points, say k, there are $k+1$ segments generated. For each segment, the procedure for estimating a single change-point is applied iteratively, which is referred to as the bisection method. A stopping criterion for the above test is based on a permutation test with the null hypothesis that there are no new change-points, where the observations within each segment are resampled to build a new series and the same change-point estimation method is applied as above. We fix a number r and this process is to be r times and the value of (3.2.5), denoted by $\hat{q}_k$, is also calculated for every permutation. For R number of permutations and significance level lying in $(0, 1)$, the approximate p-value is defined as $\frac{\text{number of } r \text{ such that } \hat{q}_k^{(r)} > \hat{q}_k}{R+1}$. The change-point τ_k is significant, given the previously estimated ones, if the approximate p-value obtained is less than the significance level, and then starts the search for the next break.

Hlavka [6] employed the concept of characteristic function as a distance measure to detect change-points. Consider X_t and Y_t that are both p dimensional and stationary. The null hypothesis in this study is that the distribution of both two sets of variables is the same and the alternative hypothesis is defined in both online and retrospective ways. Online setup is done sequentially where after each new observation, the detection criterion is calculated, where in retrospective setup, we have all the observations available. Considering pairwise detection in a multivariate setup (considering X_t's and Y_t's pairwise), the detection criteria are given by

$$D_{t,W}(\hat{\phi}_{X,t}, \hat{\phi}_{Y,t}) = \int \left|\hat{\phi}_{X,t}(u) - \hat{\phi}_{Y,t}(u)\right|^2 W(u)\, du,\ t\ = 1\ , 2\ , \ldots, \tag{3.2.6}$$

where

$$\hat{\phi}_{X,t}(u) = \frac{1}{t}\sum_{\tau=1}^{t} e^{iu^\top X_\tau} \tag{3.2.7}$$

is the empirical characteristic function from X_t. When observations arrive sequentially, the above criterion is calculated every time until the null hypothesis is rejected. Rejection is based on the computation of a threshold value as discussed in Hlavka [6]. The test criterion is modified in the retrospective setting as

$$D_{t,W,N} = \int\int \left|\hat{\phi}_{X,Y,t}(u_1, u_2) - \hat{\phi}^{(0)}_{X,Y,t}(u_1, u_2)\right|^2 W(u_1, u_2)\, du_1\, du_2,\ t\ = 1\ , 2\ , \ldots, \tag{3.2.8}$$

where

$$\begin{aligned}\hat{\phi}_{X,Y,t}(u_1, u_2) &= \frac{1}{t}\sum_{\tau=1}^{t} e^{i(u_1^\top X_\tau + u_2^\top Y_\tau)},\\ \hat{\phi}^{(0)}_{X,Y,t}(u_1, u_2) &= \frac{1}{N-t}\sum_{\tau=t+1}^{N} e^{i(u_1^\top X_\tau + u_2^\top Y_\tau)}\end{aligned} \tag{3.2.9}$$

are the empirical characteristic functions for the first t and the remaining observations, respectively. $W(.)$ denotes the weight function which is chosen to ensure computational simplicity.

Kuncheva [9] proposed a change-point detection method to overcome the drawbacks of the log-likelihood based Kullback-Leibler divergence (KL Divergence) and Hotelling's T^2 test and introduced semi-parametric log-likelihood (SPLL) detection criteria. Here, a Gaussian mixture of p components is considered with distribution $P_1(x)$ which is

$$P_1(x) = \sum_{i=1}^{p} P(i)\; p_1(x|i). \tag{3.2.10}$$

Here, $P(i)$ is the mixing coefficient and the second term in the RHS is replaced with Gaussian probability density function. Consider two consecutive time windows, TW_1 and TW_2 and let them be sampled from distributions $P_1(x)$ and $P_2(x)$ with M_1 and M_2 observations each. The idea is to check for distributional change between the two. Then, the traditional likelihood is replaced by

$$\bar{L}L = -\frac{1}{2}\sum_{X\in TW_2}(X-\mu_i)^\top \Sigma_i^{-1}(X-\mu_i) \times \log\left\{\frac{1}{(2\pi)^{N/2}\mathrm{det}(\Sigma_i)^{1/2}}\right\}. \tag{3.2.11}$$

The distribution of P_1 is estimated by using k means clustering. Here, Σ_i is assumed to be the same across all the p components. $\bar{LL}$ in (3.2.11) is proportional to negative squared Mahalanobis distances between the observations and its closest mean (μ_i that minimizes (3.2.11)). If $P_2(x)$ is the actual distribution for TW_2, then $L(TW_2|P_2) = 1$ and the likelihood ratio is given by $LLR = -\log L(TW_2|P_1)$. The proposed SPLL is given by

$$SPLL = -\frac{\bar{LL}}{M_2} \tag{3.2.12}$$

and it follows a chi-square distribution with p (number of dimensions) degrees of freedom. If TW_2 belongs to a distribution other than $P_1(x)$, the mean will deviate from p. While Hotelling's T^2 test is parametric and the test based on KL Divergence is a non-parametric criterion, the proposed method takes the semi-parametric (having both parametric and non-parametric components) approach to obtain change-points.

Chapter 4

Theoretical results

In this chapter, we establish the multivariate version of Theorem 21.1 of Billingsley [2] for the case when $X_t = g(\xi(t), \xi(t-1), \ldots)$ is a d-dimensional sequence of random variables, with g being measurable and ϵ's being m-dependent sequence of random variables, that satisfies the assumptions in Theorem 21.1 of Billingsley [2]. The following theorem helps us to perform hypothesis testing under the assumption that H_0 is true, following which we also provide a consistent estimator of the change-point under H_a.

4.1 Theorems

Theorem 4.1. *Consider a sequence $\{\xi(t)\}_{t\in\mathbb{Z}}$ of identically distributed, m-dependent, d-dimensional random variables with $\mathbb{E}[||\xi(t)||^2] < \infty$. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a strictly stationary sequence of d-dimensional random variables defined as $X_t = g(\xi(t), \xi(t-1), \ldots)$, where g is a measurable function. Denote $X_{tl} = g(\xi(t), \ldots, \xi(t-l))$. Assume*

$\mathbb{E}[\xi(t)] = 0$ and $\mathbb{E}[||X_t||^2] < \infty$ for all $t \in \mathbb{Z}$. Assume further that

$$\sum_{l \geq 1} \left(\mathbb{E}\left[||X_0 - X_{0l}||^2 \right] \right)^{1/2} < \infty.$$

Then,

$$\Sigma = \sum_{j \in \mathbb{Z}} Cov(X_0, X_j)$$

converges absolutely componentwise and if Σ is positive-definite, then

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{[Nt]} X_i \xrightarrow{D[0,1]} W_\Sigma(t).$$

Proof. Let us denote

$$S_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{[Nt]} X_i.$$

For a fixed $V \neq 0$, we denote $Y_i = V^T X_i = V^T g(\xi(i), \xi(i-1), \ldots)$ and $Y_{il} = V^T X_{il} = V^T g(\xi(i), \ldots, \xi(i-l))$, where $X_{il} = g(\xi(i), \ldots, \xi(i-l))$. Next, we have by the assumption above, that

$$\begin{aligned}
\sum_{l \geq 1} \left(\mathbb{E}\left[(Y_i - Y_{il})^2 \right] \right)^{1/2} &= \sum_{l \geq 1} \left(\mathbb{E}\left[(V^T (X_i - X_{il}))^2 \right] \right)^{1/2} \\
&= ||V||_\infty \sum_{l \geq 1} \left(\mathbb{E}\left[||X_i - X_{il}||^2 \right] \right)^{1/2} < \infty
\end{aligned}$$

and $\mathbb{E}[V^T X_i] = 0$. Hence Y_i satisfies the assumptions in Theorem 21.1 in Billingsley

[2]. Hence for $0 \leq s \leq t \leq 1$,

$$S_N(s,t) = \frac{1}{\sqrt{N}} \sum_{i=[Ns]+1}^{[Nt]} Y_i \xrightarrow{D} N(0, \sigma_Y^2(t-s)),$$

where $\sigma_Y^2 = \sum_{j \in \mathbb{Z}} \mathbb{E}[Y_0 Y_j] = \sum_{j \in \mathbb{Z}} V^T Cov(X_0, X_j) V$. (We have assumed that the matrix inside is positive definite, which implies that $\sigma_Y(t-s) > 0$ for $V \neq 0$.) Since we can choose V arbitrarily and by using similar arguments, we will have the same result. Hence by Cramer-Wald device 2.2, we have

$$\frac{1}{\sqrt{N}} \sum_{i=[Ns]+1}^{[Nt]} X_i \xrightarrow{D} W_\Sigma(t-s),$$

where $\Sigma = \sum_{j \in \mathbb{Z}} Cov(X_0, X_j)$ and W_Σ is Wiener process with Σ being the covariance matrix. For a fixed $0 < s < t$, we consider the random variables

$$\hat{S}_N(s,t) = \frac{1}{\sqrt{N}} \left(\sum_{i=[Ns]+1}^{[Nt]} X_{i([Ns]+1-i)} \right)$$

and

$$S_{N,m}(s) = \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{[Ns]-m} X_i \right)$$

for some large N. Therefore, $\hat{S}_N(s,t)$ and $\hat{S}_{N,m}(s)$ are independent random variables.

Now, upon using Markov inequality and Lemma 2.8, we get

$$\begin{aligned}
\mathbb{P}\left(||\hat{S}_N(s,t) - S_N(s,t)|| > \delta\right) &\leq \frac{1}{\delta^2}\mathbb{E}\left[||\hat{S}_N(s,t) - S_N(s,t)||^2\right] \\
&\leq \frac{1}{N\delta^2}\left(\sum_{i=[Ns]+1}^{[Nt]}\left(\mathbb{E}\left[||X_{i([Ns]+1-i)} - X_i||^2\right]\right)^{1/2}\right)^2 \\
&\leq \frac{1}{N\delta^2}\left(\sum_{i=[Ns]+1}^{[Nt]}\left(\mathbb{E}\left[||X_{1([Ns]+1-i)} - X_1||^2\right]\right)^{1/2}\right)^2 \\
&\leq \frac{1}{N\delta^2}\left(\sum_{l\geq 0}\left(\mathbb{E}\left[||X_{1l} - X_1||^2\right]\right)^{1/2}\right)^2 \xrightarrow{N\to\infty} 0,
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left(||S_N(t) - S_{N,m}(t)|| > \delta\right) &\leq \frac{1}{N\delta^2}\mathbb{E}\left[||S_N(t) - S_{N,m}(t)||^2\right] \\
&\leq \frac{1}{N\delta^2}\mathbb{E}\left[\left|\left|\sum_{i=[Ns]-m+1}^{[Ns]} X_i\right|\right|^2\right] \\
&\leq \frac{1}{N\delta^2}\left(\sum_{i=[Ns]-m+1}^{[]Ns]}\mathbb{E}\left(||X_i||^2\right)^{1/2}\right)^2 \\
&= \frac{m^2}{N\delta^2}\mathbb{E}\left(||X_i||^2\right) \xrightarrow{N\to\infty} 0.
\end{aligned}$$

Hence, we have $||S_N(s) - S_{N.m}(t)|| \xrightarrow{P} 0$ and $||\hat{S}_N(s,t) - S_N(s,t)|| \xrightarrow{P} 0$. Then Lemma 2.5 implies that

$$S_{N,m}(s) = \frac{1}{\sqrt{N}}\sum_{i=1}^{[Ns]} X_i \xrightarrow{D} N(0, \Sigma s)$$

and finally by Continuous Mapping Theorem 2.4, it follows that

$$\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{[Ns]} X_i, \frac{1}{\sqrt{N}} \sum_{i=1}^{[Nt]} X_i \right) \xrightarrow{D} \left(W_\Sigma(s), W_\Sigma(t) \right).$$

Similarly, we can show that the convergence in distribution takes place for any finite dimensional $(t_1, \ldots, t_k) \in [0,1]$, and tightness will follow when we choose $V = e_i$'s where e_i is the vector with 1 in the i-th position and zeros in all other places. This yields that the coordinates form a uniformly tight sequence of random variables and by Lemma 2.7, it finally gives the uniform tightness of $\{X_i\}_{i\geq 1}$. Hence, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{[Nt]} X_i \xrightarrow{D[0,1]} W_\Sigma(t).$$

□

Then by Theorem 4.1 and Definiton 2.11, it follows that

$$\frac{1}{\sqrt{N}} \left(\sum_{i=1}^{[Nt]} X_i - t \sum_{i=1}^{N} X_i \right) \xrightarrow{D[0,1]} W_\Sigma^0(t),$$

where $W_\Sigma^0(t)$ is Brownian bridge with $Cov(W_\Sigma^0(s), W_\Sigma^0(t)) = (\min\{s,t\} - st)\Sigma$. Hence for a fixed t, we have

$$(\tilde{S}_N(t))^T \hat{\Sigma}_N^{-1} (\tilde{S}_N(t)) \xrightarrow{D[0,1]} \sum_{i=1}^{d} (W_i^0(t))^2,$$

where

$$\tilde{S}_N(t) = \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{[Nt]} X_i - t \sum_{i=1}^{N} X_i \right)$$

and $W_i^0(t)$'s are standard one-dimensional Brownian-bridges in $[0,1]$ and $\hat{\Sigma}_N \xrightarrow{P} \Sigma$, for which we will find such estimators in Theorem 4.6.

Let $f \in D[0,1]$ and consider $N_\epsilon(f) = \{g \in D[0,1] : d_0(f,g) < \epsilon\}$ and for $\forall g \in N_\epsilon(f)$. By Definition 2.6, there exists a continuous strictly increasing function $\lambda : [0,1] \to [0,1]$ such that $\lambda(0) = 0$ and $\lambda(1) = 1$ and $\sup_{t\in[0,1]} |f(t) - g(\lambda(t))| < 3/2\epsilon$ and this implies that $|\sup_{t\in[0,1]} f(t) - \sup_{t\in[0,1]} g(t)| \leq 3/2\epsilon$. It is clear that $(\tilde{S}_N(t))^T \hat{\Sigma}_N^{-1} (\tilde{S}_N(t)) \in D[0,1]$. Hence, the function $\sup_{t\in[0,1]}(.)$ is a continuous functional on $D[0,1]$ and so applying Lemma 2.4, we have

$$\sup_{t\in[0,1]} (S_N(t))^T \hat{\Sigma}_N^{-1} (S_N(t)) \xrightarrow{D} \sup_{t\in[0,1]} \sum_{i=1}^{d} (W_i^0(t))^2.$$

Here, we consider the linear process defined by

$$X_t = \sum_{k\geq 0} C_k \xi(t-k), \tag{4.1.1}$$

where $\{\xi(t)\}_{t\in\mathbb{Z}}$ represents m-dependent identically distributed random variables. To satisfy the conditions stated in Theorem 4.1, we need

$$\sum_{l\geq 1} \mathbb{E} \left\{ \left\| \sum_{m>l} C_m \xi(-m) \right\|^2 \right\}^{1/2} < \infty.$$

Again, using Lemma 2.8, we have

$$\sum_{l\geq 1}\mathbb{E}\left\{\left|\left|\sum_{m>l}C_m\xi(-m)\right|\right|^2\right\}^{1/2}\leq\sum_{l\geq 1}\sum_{m>l}||C_m||\,\mathbb{E}\left\{||\xi(-m)||^2\right\}^{1/2}.$$

Given that $\{\xi(t)\}_{-\infty<t<\infty}$ are identically distributed and $\mathbb{E}\left\{||\xi(0)||^2\right\}^{1/2}<\infty$, if

$$\sum_{l\geq 1}\sum_{m>l}||C_m||<\infty,$$

then the condition in Theorem 4.1, namely, $\sum_{l\geq 1}\mathbb{E}\left\{||X_0-X_{0l}||^2\right\}^{1/2}<\infty$, is satisfied. This condition can be further extended to a condition based on the entries of the matrices C_i's in the following manner:

$$\begin{aligned}\sum_{l\geq 1}\sum_{m>l}||C_m||&\leq\sum_{l\geq 1}\sum_{m>l}\left(\sum_{i=1}^{d}\sum_{j=1}^{d}|(C_m)_{ij}|^2\right)^{1/2}\\&\leq\sum_{l\geq 1}\sum_{m>l}\left(\sum_{i=1}^{d}\sum_{j=1}^{d}|(C_m)_{ij}|\right)\\&\leq\sum_{i=1}^{d}\sum_{j=1}^{d}\left(\sum_{l\geq 1}\sum_{m>l}|(C_m)_{ij}|\right).\end{aligned}$$

Consequently, if we assume that the sum of the entries of the matrices C_i converges absolutely, i.e., specifically $\sum_{l\geq 1}\sum_{m>l}|(C_m)_{ij}|<\infty$ for all $1\leq i,j\leq d$, then the aforementioned condition gets satisfied. Thus, it is evident that under H_0, the X_i's constitute a strictly stationary sequence of random variables, as required in Theorem

4.1. In the next theorem, we prove that under H_a and certain assumptions, the power of the test is asymptotically 1 and we give a consistent estimator for the change-point.

Theorem 4.2. *Suppose $T^* = [k^* n]$, for some $k^* \in (0,1)$. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic sequence with $\mathbb{E}[X_0] = 0$ and $\mathbb{E}[||X_0||^2] < \infty$. Also, consider another strictly stationary and ergodic sequence $\{X_t^*\}_{t\in\mathbb{Z}}$ with $\mathbb{E}[X_0^*] \neq \mathbb{E}[X_0]$ and $\mathbb{E}[||X_0^*||^2] < \infty$. Define $Y_t = X_t$ for all $t \in [0, T^*]$ and $Y_t = X_t^*$, for all $t \in [T^*, n]$. In other words, T^* represents the change-point in the mean of the data $\{Y_t\}_{t\in\mathbb{Z}}$. Then, both*

$$\arg\max_{t\in[0,1]} \left|\left|\tilde{S}_n(t)\right|\right|$$

and

$$\arg\max_{t\in[0,1]} \frac{1}{n}\tilde{S}_n(t)\hat{\Sigma}_n^{-1}\tilde{S}_n(t)$$

are consistent estimators of the point $k^* \in (0,1)$, where $\tilde{S}_n(t) = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[nt]} Y_i - t\sum_{i=1}^{n} Y_i\right)$.

Remark 4.3. The above theorem provides us a consistent estimator of k^* and hence the estimate for the change-point will be $[k^* n]$, where n is the length of the data.

Remark 4.4. The assumption of ergodicity in the above theorem is quite reasonable from the fact that X_t's and X_t^*'s are both of the form 4.1.1, which is a measurable function of $\xi(t)$'s. And, if the observations come from the model 5.0.1 which we have used to simulate results in Chapter 5, $\xi(t)$'s are again measurable functions of $Z(t)$'s which are independent and identically distributed and hence stationary ergodic by Lemma 2.16 and hence it follows by Proposition 2.15.

Proof. By Proposition 2.14, we have for a strictly stationary ergodic sequence,

$$\sum_{i=1}^{n} \frac{X_i}{n} \xrightarrow{a.s} \mathbb{E}[X_1].$$

This implies, by using Definiton 2.1 that for a fixed $T_0 \in [0, k^*/2]$ and given $\delta > 0$, $\hat{\epsilon} = k^*\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that

$$\begin{aligned}
&\mathbb{P}\left(\left\|\frac{1}{nT_0}\sum_{k=1}^{[nT_0]} X_i\right\| < k^*\epsilon, \forall n \geq [T_0 n_0]\right) \geq 1-\delta \\
\implies &\mathbb{P}\left(\left\|\frac{1}{nT}\sum_{k=1}^{[nT]} X_i\right\| < k^*\epsilon, \forall n \geq n_0, \forall T \in [T_0, k^*]\right) \geq 1-\delta \\
\implies &\mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=1}^{[nT]} X_i\right\| < \hat{\epsilon}, \forall n \geq n_0, \forall T \in [T_0, k^*]\right) \geq 1-\delta.
\end{aligned}$$

In particular, we have

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=1}^{[nk^*]} X_i\right\| < \hat{\epsilon}, \forall n \geq n_0\right) \geq 1-\delta.$$

Hence using the previous observations, we get that

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=[nT]+1}^{[nk^*]} X_i\right\| < 2\hat{\epsilon}, \forall n \geq n_0, \forall T \in [T_0, k^*]\right) \geq 1-\delta$$

$$\implies \mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=[nT]+1}^{[nk^*]} X_i\right\| < 2\hat{\epsilon}, \forall n \geq n_0, \forall T \in [T_0, k^*]\right) \geq 1-\delta$$

$$\implies \mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=1}^{[nk^*]-[nT]} X_i\right\| < 2\hat{\epsilon}, \forall n \geq n_0, \forall T \in [T_0, k^*]\right) \geq 1-\delta$$

$$\implies \mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=1}^{[nT^{'}]} X_i\right\| < 2\hat{\epsilon}, \forall n \geq n_0, \forall T^{'} \in [0, k^*/2]\right) \geq 1-\delta$$

$$\implies \mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=1}^{[nT]} X_i\right\| < 2\hat{\epsilon}, \forall n \geq n_0, \forall T \in [0, k^*]\right) \geq 1-\delta$$

$$\implies \lim_{n\to\infty} \mathbb{P}\left(\sup_{T\in[0,k^*]}\left\|\frac{1}{n}\sum_{k=1}^{\lfloor nT\rfloor} X_i\right\| < 2\hat{\epsilon}\right) = 1$$

and similarly by the assumption that there is another strictly stationary and ergodic sequence of random variables $\{X_i^*\}_{i\geq 0}$ with $\mathbb{E}[X_i^*] = \mathbb{E}[X_0^*], \forall i \in \mathbb{Z}$ and $X_i^* = Y_i, \forall i \geq [nk^*]$.

$$\lim_{n\longrightarrow\infty} \mathbb{P}\left(\sup_{T\in[0,k^*]}\left\|\frac{1}{n}\sum_{k=[nk^*]+1}^{[nt]} X_i^* - (t-k^*)\mathbb{E}[X_0^*]\right\| > 2\hat{\epsilon}\right) = 0.$$

We thus obtain

$$\mathbb{P}\left(\sup_{t\in[0,1]}\left\|\frac{1}{\sqrt{n}}\tilde{S}_n(t) - g(t)\right\| > \epsilon\right) \xrightarrow{n\longrightarrow\infty} 0,$$

where

$$\tilde{S}_n(t) = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{[nt]} Y_i - t \sum_{i=1}^{n} Y_i \right)$$

and

$$g(t) = \begin{cases} t(1-k^*)(\mathbb{E}[X_0] - \mathbb{E}[X_0^*]), & \text{if } t \in [0, k^*] \\ k^*(1-t)(\mathbb{E}[X_0] - \mathbb{E}[X_0^*]), & \text{if } t \in [k^*, 1]. \end{cases}$$

Therefore $\sup_{t\in[0,1]} \left| \frac{1}{\sqrt{n}} \tilde{S}_n(t) - g(t) \right| \xrightarrow[P]{N \to \infty} 0$. Hence, for a given $\epsilon > 0$, $\exists\ \delta > 0, N_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{t\in[0,1]} \left|\left| \frac{1}{\sqrt{n}} \tilde{S}_n(t) - g(t) \right|\right| < \delta \right) > 1 - \epsilon, \forall n \geq N_0.$$

Hence, $\forall n \geq N_0$ with probability at least $1 - \epsilon$,

$$\sqrt{(k^*)^2 - 2\delta} < \arg\max_{t\in[0,k^*]} \left|\left| \frac{1}{\sqrt{n}} \tilde{S}_n(t) \right|\right|^2 \leq k^*$$

and similarly,

$$k^* \leq \arg\max_{t\in[0,k^*]} \left|\left| \frac{1}{\sqrt{n}} \tilde{S}_n(t) \right|\right|^2 < \sqrt{(k^*)^2 + 2\delta}.$$

Thus, for $n \geq N_0$, we have an injective function $\hat{\Delta}(\delta)$ dependent only on k^* such that

$$\mathbb{P}\left(\left| \arg\max_{t\in[0,k^*]} \left|\left| \frac{1}{\sqrt{n}} \tilde{S}_n(t) \right|\right|^2 - k^* \right| < \hat{\Delta}(\delta) \right) > 1 - \epsilon,$$

and hence

$$\arg\max_{t\in[0,1]} \left|\left| \tilde{S}_n(t) \right|\right|$$

and similarly

$$\arg\max_{t\in[0,1]} \frac{1}{n} \tilde{S}_n(t) \hat{\Sigma}_n^{-1} \tilde{S}_n(t)$$

are consistent estimators of $k^* \in [0,1]$ using Definition 2.1. $\square$

For a sequence $\{X_1, \ldots, X_N\}$, where X_t is a stationary time series with $\mathbb{E}[X_t] = \mu$ and covariance matrices $\mathbb{E}[X_{t+h}X_t^T] - \mu\mu^T = \Gamma(h)$ have absolutely summable components. Under this condition, X_t has a continuous spectral density given by

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-ih\lambda}, \quad -\pi \leq \lambda \leq \pi.$$

For $\omega_j = \frac{2\pi j}{N}$, where $-[(N-1)/2] \leq j \leq [N/2]$, the discrete Fourier transform of $\{X_1, \ldots, X_N\}$ is defined by

$$W_N(\omega_j) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n e^{in\omega_j}$$

and let $I_N(\omega_j)$ denote the periodogram of $\{X_1, \ldots, X_N\}$, defined by $I_N(\omega_j) = W(\omega_j)W(\omega_j)^*$. The periodogram is a commonly used tool in signal processing and spectral analysis to estimate the spectral density of a time series. Given a stationary time series $\{X_t\}$, the periodogram $I_N(\omega_j)$ at frequency $\omega_j = 2\pi j/N$ is defined as

$$I_N(\omega_j) = \frac{1}{N} \left(\sum_{n=1}^{N} X_n e^{in\omega_j} \right) \left(\sum_{n=1}^{N} X_n e^{in\omega_j} \right)^*,$$

where N is the length of the time series. The definition of $I_N(\omega)$ is then extended to $\omega \in [-\pi, \pi]$ by the following relation

$$I_N(\omega) = \begin{cases} I_N(g(N,\omega)), & \text{if } \omega \geq 0, \\ I_N^*(g(N,-\omega)), & \text{if } \omega < 0, \end{cases}$$

where $g(N,\omega)$, where $\omega \in [0,\pi]$, is the nearest multiple of $2\pi/N$ to ω.

In the following theorem, we estimate the spectral density $f(\omega)$, where $\omega \in [-\pi,\pi]$. We consider an estimator for $f(\omega)$, as $\hat{f}(\omega_j) = (2\pi)^{-1}\sum_{|k|\le h_N} K_N(k)I_N(\omega_{j+k})$, where $\{h_N\}$ is a sequence of positive integers and $\{K_N(.)\}$ is a sequence of weight functions. We also impose the following conditions on $\{h_N\}$ and $\{K_N(.)\}$:

$$\begin{aligned}
& h_N \to \infty \text{ and } h_N/\sqrt{N} \to 0 \text{ as } N \to \infty, \\
& K_N(k) = K_N(-k), \qquad K_N(k) \ge 0, \text{ for all } k \\
& \sum_{|k|\le h_N} K_N(k) = 1 \\
\text{and } & \sum_{|k|\le h_N} K_N^2(k) \xrightarrow{N\to\infty} 0.
\end{aligned} \tag{4.1.2}$$

Remark 4.5. The above assumptions are satisfied by the Simple moving average kernel

$$W_N(k) = \frac{1}{(2h_N+1)}.\mathbb{I}_{\{|k|\le h_N\}}, \tag{4.1.3}$$

where h_N is the chosen sequence of bandwidth.

Theorem 4.6. *Under the assumptions on X_t based on Theorem 4.1 and the weight functions $K_N(k)$ or under H_0, $\hat{f}(\omega) = (2\pi)^{-1}\sum_{|k|\le h_N} K_N(k)I_N(g(N,\omega)+\omega_k)$ is a consistent estimator of $f(\omega)$, for $-\pi \le \omega \le \pi$. Consequently, there exists a consistent estimator $\hat{\Sigma}_N$ for Σ.*

Proof. To prove this, we follow similar steps as in Section 10.3 in Brockwell [3] where they prove the univariate version for linear processes with independent innovations.

We similarly define the discrete Fourier transform of $\{\xi(1), \ldots, \xi(N)\}$ as

$$W_{N,\xi}(\lambda) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \xi(n) e^{in\lambda}.$$

For $\omega_j = \frac{2\pi j}{N}$, where $-[(N-1)/2] \leq j \leq [N/2]$, $I_{N,\xi}(\omega_j)$ denotes the periodogram of $\{\xi(1), \ldots, \xi(N)\}$, defined by $I_{N,\xi}(\omega_j) = W_{N,\xi}(\omega_j) W_{N,\xi}(\omega_j)^*$. Thus,

$$(I_{N,\xi}(\omega_j))_{pq} = \frac{1}{N} \left(\sum_{n=1}^{N} \xi_p(n) e^{in\omega_j} \right)^* \left(\sum_{n=1}^{N} \xi_q(n) e^{in\omega_j} \right).$$

$$\mathbb{E}\left[(I_{N,\xi}(\omega_j))_{pq} (I_{N,\xi}(\omega_k))_{rs} \right] = \frac{1}{N^2} \sum_{s,t,u,v=1}^{N} \mathbb{E}\left[\xi_p(s)\xi_q(t)\xi_r(u)\xi_s(v) e^{i(t-s)\omega_j} e^{i(v-u)\omega_k} \right]$$

The last sum can be decomposed into four types of sums (since under H_0, we have assumed that $\mathbb{E}[\xi(t)] = 0$) as follows:

$$= \sum\sum\sum\sum \mathbb{E}\left[\xi_p(s)\xi_q(t)\xi_r(u)\xi_s(v)\right] e^{i(t-s)\omega_j} e^{i(v-u)\omega_k} \quad \text{(i)}$$

$$+ \sum\sum\sum\sum \mathbb{E}\left[\xi_p(s)\xi_q(t)\right]\mathbb{E}[\xi_r(u)\xi_s(v)] e^{i(t-s)\omega_j} e^{i(v-u)\omega_k} \quad \text{(ii)}$$

$$+ \sum\sum\sum\sum \mathbb{E}\left[\xi_p(s)\xi_q(u)\right]\mathbb{E}[\xi_r(t)\xi_s(v)] e^{i(t-s)\omega_j} e^{i(v-u)\omega_k} \quad \text{(iii)}$$

$$+ \sum\sum\sum\sum \mathbb{E}\left[\xi_p(s)\xi_q(v)\right]\mathbb{E}[\xi_r(t)\xi_s(v)] e^{i(t-s)\omega_j} e^{i(v-u)\omega_k} \quad \text{(iv).}$$

For the type-i sum, since the $\xi(i)$'s are m-dependent random variables, it should be of $O(N)$. For the type-iii sum, we consider the following expressions,

$$
\begin{aligned}
&\sum_{t=m+1}^{N-m}\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i\omega_j t}e^{i\omega_k(t+\alpha)}\\
&=\sum_{t=m+1}^{N-m}\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i(\omega_j+\omega_k)t}e^{i\omega_k\alpha}\\
&=\sum_{t=1}^{N}\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i(\omega_j+\omega_k)t}e^{i\omega_k\alpha}\\
&\qquad-\sum_{t=1}^{m}\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i(\omega_j+\omega_k)t}e^{i\omega_k\alpha}\\
&\qquad-\sum_{t=N-m+1}^{N}\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i(\omega_j+\omega_k)t}e^{i\omega_k\alpha}\\
&=\left(\sum_{t=1}^{N}e^{i(\omega_j+\omega_k)t}\right)\left(\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i\omega_k\alpha}\right)\\
&\qquad-\sum_{t=1}^{m}\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i(\omega_j+\omega_k)t}e^{i\omega_k\alpha}\\
&\qquad-\sum_{t=N-m+1}^{N}\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i(\omega_j+\omega_k)t}e^{i\omega_k\alpha}\\
&=\left(\sum_{t=1}^{N}e^{i(\omega_j+\omega_k)t}\right)\left(\sum_{\alpha=-m}^{m}\mathbb{E}\left[\xi_p(t)\xi_q(t+\alpha)\right]e^{i\omega_k\alpha}\right)+O(1).
\end{aligned}
$$

In the third equality of the above expression, we have used the fact that $\xi(n)'s$ is a stationary sequence of random variables and so we can decompose it into products which do not depend on t. And the remaining part of the expression in the third equality is of $O(1)$ since there are m terms, where m is merely a constant and hence again by stationary property of $\xi(n)'s$, we have the rest to be of constant order. This means the following:

$$\text{If } \omega_j = \omega_k = 0, \text{ or } \omega_j = \omega_k = \pi, \text{ then } = O(N);$$
$$\text{Otherwise, then } = O(1).$$

Therefore, the following is true for type-iii sum:

$$\text{If } \omega_j = \omega_k = 0, \text{ or } \omega_j = \omega_k = \pi, \text{ then } = O(N^2);$$
$$\text{Otherwise, then } = O(1).$$

Similar arguments will work for the type-iv sum as well and will give the same result. We thus have

$$\begin{aligned}
&\text{Cov}\left((I_{N,\xi}(\omega_j))_{pq}, (I_{N,\xi}(\omega_k))_{rs}\right) \\
&= \mathbb{E}\left(\left(I_{N,\xi}(\omega_j)\right)_{pq}\left(I_{N,\xi}(\omega_k)\right)_{rs}\right) - \mathbb{E}\left(I_{N,\xi}(\omega_j)\right)_{pq}\mathbb{E}\left(I_{N,\xi}(\omega_j)\right)_{pq} \\
&= \mathbb{E}\left(\left(I_{N,\xi}(\omega_j)\right)_{pq}\left(I_{N,\xi}(\omega_k)\right)_{rs}\right) \\
&\qquad\qquad - \frac{1}{N^2}\sum_{s,t,u,v=1}^{N} \mathbb{E}\left[\xi_p(s)\xi_q(t)\right]\mathbb{E}[\xi_r(u)\xi_s(v)]\, e^{i(t-s)\omega_j} e^{i(v-u)\omega_k}.
\end{aligned}$$

Hence combining the above orders, we get

$$\text{Cov}\left((I_{N,\xi}(\omega_j))_{pq}, (I_{N,\xi}(\omega_k))_{rs}\right) = \begin{cases} O\left(\frac{1}{N}\right), & \text{if } 0 < \omega_j \neq \omega_k < \pi, \\ O(1), & \text{if } 0 \leq \omega_j = \omega_k \leq \pi. \end{cases} \tag{4.1.4}$$

Likewise, the discrete Fourier transform of $\{X_1, \ldots, X_N\}$ is defined as

$$
\begin{aligned}
W_{N,X}(\lambda) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n e^{in\lambda} \\
&= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} e^{in\lambda} \left\{ \sum_{k \geq 0} C_k \xi(N-k) \right\} \\
&= \sum_{k \geq 0} C_k e^{ik\lambda} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \xi(n) e^{in\lambda} + \frac{1}{\sqrt{N}} \sum_{k \geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda),
\end{aligned}
$$

where

$$
\left(R_{k,N}(\lambda)\right)_p = \mathbb{E}\left[-\sum_{n=1}^{N} \xi_p(n) e^{in\lambda} - \sum_{n=-k+1}^{N-k} \xi_p(n) e^{in\lambda} \right].
$$

Let $(I_{N,X}(\lambda))$ denote the periodogram of $\{X_1, \ldots, X_N\}$, defined by

$I_{N,X}(\lambda) = W_{N,X}(\lambda)W_{N,X}(\lambda)^*$. Thus

$$\begin{aligned}
I_{N,X}(\lambda) &= \frac{1}{N}\left(\sum_{n=1}^{N} X_n e^{in\lambda}\right)\left(\sum_{n=1}^{N} X_n e^{in\lambda}\right)^* \\
&= \left(\sum_{k\geq 0} C_k e^{ik\lambda}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\lambda} + \frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right) \\
&\qquad \times \left(\sum_{k\geq 0} C_k e^{ik\lambda}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\lambda} + \frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)^* \\
&= \left(\sum_{k\geq 0} C_k e^{ik\lambda}\right) I_{N,\xi}(\lambda)\left(\sum_{k\geq 0} C_k e^{ik\lambda}\right)^* \\
&\qquad + \left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)^* \\
&\qquad + \left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)\left(\sum_{k\geq 0} C_k e^{ik\lambda}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\lambda}\right)^* \\
&\qquad + \left(\sum_{k\geq 0} C_k e^{ik\lambda}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\lambda}\right)\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)^* \\
&= \left(\sum_{k\geq 0} C_k e^{ik\lambda}\right) I_{N,\xi}(\lambda)\left(\sum_{k\geq 0} C_k e^{ik\lambda}\right)^* + R_N(\lambda), \qquad (4.1.5)
\end{aligned}$$

where

$$\begin{aligned}
R_N(\lambda) &= \left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)^* \\
&\qquad + \left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)\left(\sum_{k\geq 0} C_k e^{ik\lambda}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\lambda}\right)^* \\
&\qquad + \left(\sum_{k\geq 0} C_k e^{ik\lambda}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\lambda}\right)\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda}R_{k,N}(\lambda)\right)^*.
\end{aligned}$$

By using Lemma 2.8,

$$\begin{aligned}
&\max_{\lambda} \frac{1}{\sqrt{2\pi N}} \mathbb{E}\left[\left(\left(\sum_{j=-\infty}^{\infty} C_j e^{ij\lambda} R_{j,N}(\lambda)\right)^* \left(\sum_{j=-\infty}^{\infty} C_j e^{ij\lambda} R_{j,N}(\lambda)\right)\right)^2\right]^{1/4} \\
&\qquad \le \frac{1}{\sqrt{2\pi N}} \max_{\lambda} \sum_{j=-\infty}^{\infty} \mathbb{E}\left[\left|\left|(C_j e^{ij\lambda} R_{j,N}(\lambda))\right|\right|^4\right]^{1/4} \\
&\qquad \le \frac{1}{\sqrt{2\pi N}} \max_{\lambda} \sum_{j=-\infty}^{\infty} ||C_j|| \, \mathbb{E}\left[||(R_{j,N}(\lambda))||^4\right]^{1/4}
\end{aligned}$$

and

$$\mathbb{E}\left[||R_{k,N}(\lambda)||^4\right] = \mathbb{E}\left[\left(\sum_{m,n\in I} \xi_p\left(m\right) \xi_p\left(n\right) e^{i(m-n)\lambda}\right)^2\right] = O\left(\min(|k|^2, |N|^2)\right),$$

where $I = \{1, \ldots, N\} \Delta \{-k+1, \ldots, N-k\}$. Therefore by the assumption that

$$\sum_{l\ge 1} \sum_{m>l} ||C_m|| < \infty,$$

we have

$$\begin{aligned}
&\max_{\lambda} \mathbb{E}\left[\left(\left(\frac{1}{\sqrt{2\pi N}} \sum_{j=-\infty}^{\infty} C_j e^{ij\lambda} R_{j,N}(\lambda)\right)^* \left(\frac{1}{\sqrt{2\pi N}} \sum_{j=-\infty}^{\infty} C_j e^{ij\lambda} R_{j,N}(\lambda)\right)\right)^2\right] \\
&\le O\left(\frac{1}{N^2}\right).
\end{aligned} \tag{4.1.6}$$

and

$$\begin{aligned}\mathbb{E}\left[\left|C(e^{i\omega_k})W_{N,\xi}(\omega_k)r_N^*(\omega_k)\right|_{ij}^2\right] &= \mathbb{E}\left[\left|C(e^{i\omega_k})W_{N,\xi}(\omega_k)\right|_i^2 \left|r_N^*(\omega_k)\right|_j^2\right] \\ &\leq \mathbb{E}\left[\left|C(e^{i\omega_k})W_{N,\xi}(\omega_k)\right|_i^4\right]^{1/2} \mathbb{E}\left[\left|r_N(\omega_k)\right|_j^4\right]^{1/2} \\ &\leq \left|\left|C(e^{i\omega_k})\right|\right|^2 \mathbb{E}\left[\left|W_{N,\xi}(\omega_k)\right|_i^4\right]^{1/2} \mathbb{E}\left[\left|r_N(\omega_k)\right|_j^4\right]^{1/2},\end{aligned}$$

where $r_N(\omega_k) = \frac{1}{\sqrt{N}}\sum_{j=-\infty}^{\infty} C_j e^{ij\omega_k} R_{j,N}(\omega_k)$ for $\omega_k = \frac{2\pi k}{N} \in [0,\pi]$. As we already have seen that

$$\begin{aligned}&\mathrm{Cov}\left((I_N(\omega_k))_{pq}, (I_N(\omega_k))_{rs}\right) \\ &= \mathbb{E}\left[\left((W_{N,\xi}(\omega_k))_p\overline{(W_{N,\xi}(\omega_k))_q}\right)\left((W_{N,\xi}(\omega_k))_r\overline{(W_{N,\xi}(\omega_k))_s}\right)\right] \\ &\qquad - \mathbb{E}\left[\left((W_{N,\xi}(\omega_k))_p\overline{(W_{N,\xi}(\omega_k))_q}\right)\right]\mathbb{E}\left[\left((W_{N,\xi}(\omega_k))_r\overline{(W_{N,\xi}(\omega_k))_s}\right)\right] \\ &= O\left(\frac{1}{N}\right) + O(1)\end{aligned}$$

and

$$\mathbb{E}\left[\left((W_{N,\xi}(\omega_k))_p\overline{(W_{N,\xi}(\omega_k))_q}\right)\right]\mathbb{E}\left[\left((W_{N,\xi}(\omega_k))_r\overline{(W_{N,\xi}(\omega_k))_s}\right)\right] = O(1)$$

since the sequence $\{\xi\,(n)\}_{-\infty<n<\infty}$ are m-dependent random variables, we have

$$\mathbb{E}\left[\left((W_{N,\xi}(\omega_k))_p\overline{(W_{N,\xi}(\omega_k))_q}\right)\right] = \frac{1}{2\pi N}\sum_{j,k=1}^{N}\mathbb{E}\left[\xi_p\,(j)\,\xi_q\,(k)\right]e^{i(j-k)\omega_k} = O\,(1)\,.$$

Hence,

$$\mathbb{E}\left[\left(W_p(\omega_k)\overline{W_q(\omega_k)}\right)\left(W_r(\omega_k)\overline{W_s(\omega_k)}\right)\right] = O\left(\frac{1}{N}\right) + O(1),$$

which implies that

$$\max_{\omega_k \in [0,\pi]} \mathbb{E}\left[|W(\omega_k)|_i^4\right] = O\left(\frac{1}{N}\right) + O(1),$$

$$\begin{aligned}\max_{\omega_k \in [0,\pi]} \mathbb{E}\left[\left|C(e^{i\omega_k})W(\omega_k)r_N^*(\omega_k)\right|_{ij}^2\right] &\leq \left|\left|C(e^{i\omega_k})\right|\right|^2 \left(O\left(\frac{1}{N}\right) + O\left(1\right)\right)^{1/2} \left(O\left(\frac{1}{N^2}\right)\right)^{1/2} \\ &= O\left(\frac{1}{N}\right), \end{aligned} \tag{4.1.7}$$

and

$$\begin{aligned} &\max_\lambda \mathbb{E}\left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)^*\right|_{ij}^2 \\ &= \max_\lambda \mathbb{E}\left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)_i \overline{\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)_j}\right|^2 \\ &= \max_\lambda \mathbb{E}\left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)_i\right|^2 \left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)_j\right|^2 \\ &\leq \max_\lambda \mathbb{E}\left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)\right|^2 \left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)\right|^2 \\ &\leq \max_\lambda \left(\mathbb{E}\left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)\right|^4 \mathbb{E}\left|\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ik\lambda} R_{k,N}(\lambda)\right)\right|^4\right)^{1/2} \\ &= O\left(\frac{1}{N^2}\right). \end{aligned} \tag{4.1.8}$$

This is true from the observation obtained in (4.1.6). Finally by Lemma 2.8, we have

$$
\begin{aligned}
&\max_{\omega_k\in[0,\pi]} \left(\mathbb{E}\left[|R_N(\omega_k)|^2_{ij}\right]\right)^{1/2} \\
&\leq \max_{\omega_k\in[0,\pi]} \mathbb{E}\left(\left[\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ij\omega_k} R_{j,N}(\omega_k)\right)\left(\frac{1}{\sqrt{N}}\sum_{k\geq 0} C_k e^{ij\omega_k} R_{j,N}(\omega_k)\right)^*\right]^2_{ij}\right)^{1/2} \\
&+ \max_{\omega_k\in[0,\pi]} \left(\mathbb{E}\left[\left(\sum_{k\geq 0} C_k e^{ij\omega_k} R_{j,N}(\omega_k)\right)\left(\sum_{k\geq 0} C_k e^{ik\omega_k}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\omega_k}\right)^*\right]^2_{ij}\right)^{1/2} \\
&+ \max_{\omega_k\in[0,\pi]} \left(\mathbb{E}\left[\left(\sum_{k\geq 0} C_k e^{ik\omega_k}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\xi(n)e^{in\omega_k}\right)\left(\sum_{k\geq 0} C_k e^{ij\omega_k} R_{j,N}(\omega_k)\right)^*\right]^2_{ij}\right)^{1/2} \\
&\leq O\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}
$$

Hence,

$$
\max_{\omega_k\in[0,\pi]} \mathbb{E}\left[|R_N(\omega_k)|^2_{ij}\right] \leq O\left(\frac{1}{N}\right). \tag{4.1.9}
$$

Therefore from (4.1.5), we have

$$
\begin{aligned}
(I_{N,X}(\omega_k))_{ij} &= \left(C(e^{i\omega_k})I_{N,Z}(\omega_k)C(e^{i\omega_k})^*\right)_{ij} + (R_N(\omega_k))_{ij} \\
&= \sum_{1\leq m,n\leq d}\left[\left(C(e^{i\omega_k})\right)_{im}(I_{N,Z}(\omega_k))_{mn}\left(C(e^{i\omega_k})\right)_{nj}\right] + (R_N(\omega_k))_{ij}.
\end{aligned}
$$

which implies that

$$\begin{aligned}
&\text{Cov}\left((I_{N,X}(\omega_j))_{pq}, (I_{N,X}(\omega_k))_{rs}\right) \\
&= \text{Cov}\Bigg(\sum_{1\le m,n\le d} \left(C(e^{i\omega_j})\right)_{pm} (I_{N,\xi}(\omega_j))_{mn} \left(C(e^{i\omega_j})\right)_{nq}, \\
&\qquad\qquad \sum_{1\le m,n\le d} \left(C(e^{i\omega_k})\right)_{rm} (I_{N,\xi}(\omega_k))_{mn} \left(C(e^{i\omega_k})\right)_{ns} \Bigg) \\
&\qquad\qquad + \text{Cov}\left(\sum_{1\le m,n\le d} \left(C(e^{i\omega_j})\right)_{pm} (I_{N,\xi}(\omega_j))_{mn} \left(C(e^{i\omega_j})\right)_{nq}, (R_N(\omega_k))_{rs} \right) \\
&\qquad\qquad + \text{Cov}\left((R_N(\omega_j))_{pq}, \sum_{1\le m,n\le d} \left(C(e^{i\omega_k})\right)_{rm} (I_{N,\xi}(\omega_k))_{mn} \left(C(e^{i\omega_k})\right)_{ns} \right) \\
&\qquad\qquad + \text{Cov}\left((R_N(\omega_j))_{pq}, (R_N(\omega_k))_{rs}\right).
\end{aligned}$$

Hence from the relations in (4.1.4), (4.1.6), (4.1.7), (4.1.8), (4.1.9) and using Cov(. , .) function and Cauchy-Schwarz inequality, we obtain

$$\text{Cov}\left((I_{N,X}(\omega_j))_{pq}, (I_{N,X}(\omega_k))_{rs}\right) = \begin{cases} O\left(\frac{1}{\sqrt{N}}\right) + O(1), & \text{if } 0 \le \omega_j = \omega_k \le \pi \\ O\left(\frac{1}{\sqrt{N}}\right), & \text{if otherwise.} \end{cases}$$

By the assumptions on h_N based on (4.1.2) that

$$\frac{h_N}{\sqrt{N}} \xrightarrow{N\to\infty} 0,$$

we have

$$\max_{|k| \leq h_N} |g(N,\omega) + \omega_k - \omega| \xrightarrow{N \hookrightarrow \infty} 0.$$

This implies that, by continuity of $f_{ij}(.)$ on the compact set $[0,\pi]$, and hence by uniform continuity, we have

$$\max_{|k| \leq h_N} |f_{ij}\left(g(N,\omega) + \omega_k\right) - f_{ij}\left(\omega\right)| \xrightarrow{N \hookrightarrow \infty} 0.$$

As

$$\begin{aligned} \left|\mathbb{E}\hat{f}_{ij}(\omega) - f_{ij}(\omega)\right| = \Bigg| \sum_{|k| \leq h_N} & K_N(k)\big[(2\pi)^{-1}\mathbb{E}\left(I_{N,X}\left(g(N,\omega) + \omega_k\right)_{ij}\right) \\ & - f_{ij}\left(g(N,\omega) + \omega_k\right) + f_{ij}\left(g(N,\omega) + \omega_k\right) - f_{ij}\left(\omega\right)\big]\Bigg|, \end{aligned}$$

by Proposition 10.3.1 in Brockwell [3], for stationary sequence X_t's, we have

$$\max_{|k| \leq h_N} \left|(2\pi)^{-1}\mathbb{E}\left(I_{N,X}\left(g(N,\omega) + \omega_k\right)_{ij}\right) - f_{ij}\left(g(N,\omega) + \omega_k\right)\right| \xrightarrow{N \to \infty} 0.$$

Therefore, we have $\mathbb{E}\hat{f}(\omega) \xrightarrow{N\longrightarrow\infty} f(\omega)$. Hence for $\omega \in (0,\pi)$,

$$\begin{aligned}
\operatorname{Var}\left(\hat{f}_{pq}(\omega)\right) &= (2\pi)^{-2} \sum_{|j|\le h_N} K_N^2(j) \left((2\pi)^2 \operatorname{Var}\left(\left(I_{N,X}(g(N,\omega)+\omega_j)\right)_{pq}\right)\right) \\
&+ (2\pi)^{-2} \sum_{|j|\le h_N} \sum_{|k|\le h_N, k\ne j} K_N(j)K_N(k) \\
&\quad \times \left((2\pi)^2 \operatorname{Cov}\left(\left(I_{N,X}(g(N,\omega)+\omega_j)\right)_{pq}, 0\left(I_{N,X}(g(N,\omega)+\omega_k)\right)_{pq}\right)\right) \\
&= \left(\sum_{|j|\le h_N} K_N^2(j)\right) O(1) + (2h_N+1)\left(\sum_{|j|\le h_N} K_N^2(j)\right) O\left(\frac{1}{\sqrt{N}}\right) \\
&= \left(\sum_{|j|\le h_N} K_N^2(j)\right) O(1) + o\left(\sum_{|j|\le h_N} K_N^2(j)\right).
\end{aligned}$$

This implies that

$$\operatorname{Var}\left(\hat{f}_{pq}(\omega)\right) \xrightarrow{N\longrightarrow\infty} 0, \text{ when } 0<\omega<\pi.$$

For $\omega = 0$, we have

$$\begin{aligned}
\hat{f}(0) &= (2\pi)^{-1} \sum_{|k|\le h_N} K_N(k) I_N(\omega_k) \\
&= (2\pi)^{-1} \sum_{0\le k\le h_N} K_N(k) I_N(\omega_k) + (2\pi)^{-1} \sum_{-h_N\le k<0} K_N(k) I_N(\omega_k)
\end{aligned}$$

and let us denote the sums as $\hat{f}_1(0)$ and $\hat{f}_2(0)$, respectively. Hence,

$$\text{Var}\left(\hat{f}_{1pq}(0)\right) = (2\pi)^{-2} \sum_{0\le j\le h_N} K_N^2(j)\left((2\pi)^2\text{Var}\left(\left(I_{N,X}(\omega_j)\right)_{pq}\right)\right)$$
$$+ (2\pi)^{-2} \sum_{0\le j\le h_N} \sum_{0\le k\le h_N, k\ne j} K_N(j)K_N(k)\left((2\pi)^2\text{Cov}\left(\left(I_{N,X}(\omega_j)\right)_{pq}, \left(I_{N,X}(\omega_k)\right)_{pq}\right)\right)$$
$$\le \left(\sum_{|j|\le h_N} K_N^2(j)\right) O(1) + (2h_N+1)\left(\sum_{|j|\le h_N} K_N^2(j)\right) O\left(\frac{1}{\sqrt{N}}\right)$$
$$= \left(\sum_{|j|\le h_N} K_N^2(j)\right) O(1) + o\left(\sum_{|j|\le h_N} K_N^2(j)\right) \xrightarrow{N\to\infty} 0$$

and similarly, $\text{Var}\left(\hat{f}_{2pq}(0)\right) \xrightarrow{N\to\infty} 0$. Since $\hat{f}_{pq}(0) = \hat{f}_{1pq}(0) + \hat{f}_{2pq}(0)$,

$$\text{Var}(\hat{f}_{pq}(0)) \le \text{Var}(\hat{f}_{1pq}(0)) + \text{Var}(\hat{f}_{2pq}(0)) + 2\sqrt{\text{Var}(\hat{f}_{1pq}(0))\text{Var}(\hat{f}_{2pq}(0))}$$

by Cauchy-Schwarz inequality. Therefore,

$$\text{Var}\hat{f}_{pq}(0) \xrightarrow{N\to\infty} 0,$$

and hence,

$$\mathbb{E}\left|\hat{f}_{pq}(0) - f_{pq}(0)\right|^2 = \text{Var}\hat{f}_{pq}(0) + \left|\mathbb{E}\hat{f}_{pq}(0) - f_{pq}(0)\right|^2 \xrightarrow{N\to\infty} 0,$$

as required. This implies that $f(\hat{0})$ is a consistent estimator of $f(0)$. Hence, proved.

□

Theorem 4.2 provides a consistent estimator for the change-point when there is a single change-point. This can be extended to multiple change-points case, say, the number of change-points is k, by assuming that the distance between any two change-points is at least $\left[\frac{N}{k}\right]$. We can then apply the result in Theorem 4.2 for the data from $(i-1)\left[\frac{N}{k}\right]$ to $i\left[\frac{N}{k}\right]$, for $i \in \left\{1, \ldots, \left[\frac{N}{k}\right]\right\}$.

Chapter 5

Empirical results

In this chapter, we use T to represent the length of a given data. The empirical results were obtained by utilizing the critical values as detailed in the paper by Kiefer [8] and using Simple moving average kernel with bandwidth $h_T = T^{1/4}$, which satisfies the assumptions 4.1.2. For the task of single change-point detection, the proposed method was applied to datasets spanning different time lengths, namely, 8000, and 16000, across a spectrum of m-dependence levels including 10 and 20. As expected, given the consistent nature of our estimator and its convergence to the requisite statistic discussed in Kiefer [8] for large T, we observe an improvement in the performance of the method with increasing T. Moreover, since the assumed constant-order dependence of $\xi(n)'s$ suggests a relationship, where $O(m) << O(T^{1/4})$ as obtained in the proof of Theorem 4.6, the selection of m becomes crucial in optimizing the method's efficacy relative to the dataset size T. Additionally, it is worth noting that the performance of the method may be influenced by the location of the actual change-point, given the utilization of the convergence of the average of $X_t's$ as T tends to infinity in the proof of Theorem 4.2. Another potential factor impacting method efficacy is the magnitude of the change itself, which warrants consideration in our evaluation.

For generating the data which follows the assumptions in Theorem 4.1, we generated a sequence of independent and identically distributed random variables $\{Z(t)\}_{t\in\mathbb{Z}}$, and then generated $\xi(t) = \tilde{f}(Z(t-m), \ldots, Z(t+m)), \forall\ t \in \mathbb{Z}$, with some suitable function $\tilde{f}(\cdot)$. Hence, the data we used for analysis are as follows:

$$X_t = \sum_{k\geq 0} C_k \tilde{f}(Z(t-k-m), \ldots, Z(t), \ldots, Z(t-k+m)), \tag{5.0.1}$$

where m is the dependence parameter of the $\xi(t)$'s. This is a specific model which generates such data and we have used this to obtain our simulation results.

For each combination in the tables, we have used 30 runs and took the average of the deviations of the estimates of the change-points from the actual change-point, absolute value of deviations and square of the deviations to estimate $\mathbb{E}\left[T^* - \hat{T}\right]$ (deviation), $\mathbb{E}\left|T^* - \hat{T}\right|$ (abs deviation) and $\mathbb{E}\left|T^* - \hat{T}\right|^2$ (sqd deviation), where $\hat{T}$ and T^* are the estimated and actual change-points, respectively. While using the model in (5.0.1), we used normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\begin{Bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{Bmatrix}$$

and

$$\begin{Bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{Bmatrix},$$

for generating $\{Z(t)\}_{t\in\mathbb{Z}}$ for the bi-variate and five-variate data cases, respectively, and observed the performance of the method based on two magnitude of change in the mean $(0.5, 0.2)$ and $(0.5, 1.2)$ for the bi-variate case and $(0.5, 0.2, 0.2, 0.5, 0.2)$ and $(0.5, 1.2, 0.5, 0.5, 0.5)$ for the five-variate case. The Figures 5.1 and 5.2 show the simulated bi-variate data with the actual change-point at $T/2$ and $T/5$, respectively. It is clear from the Figures that it is difficult to identify the change-points, mainly because of the spread of the data.

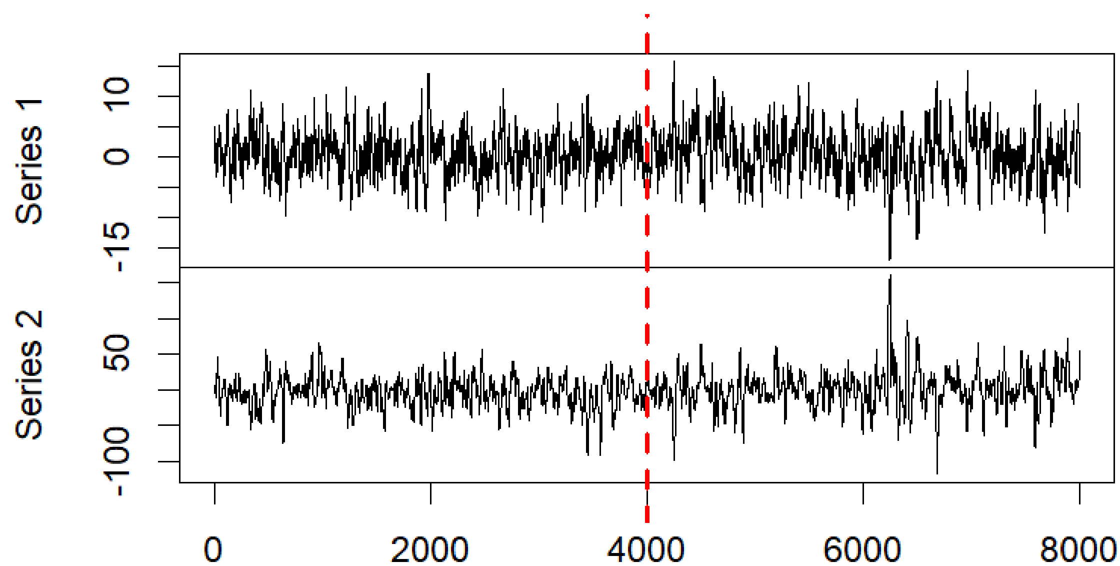

Figure 5.1: Data for the bi-variate with T = 8000 and change of mean being (0.2,0.5) at time T/2.

For each simulation, we perform the hypothesis testing, based on the results obtained in Theorem 4.1 at a significance level of 0.05 and on an average, we couldn't reject the null hypothesis 1 or 2 out of 30 simulations for each case. And once the null hypothesis gets rejected, we estimate the change-point based on the data simulated. We discuss some of the important observations from the Tables 5.1-5.4.

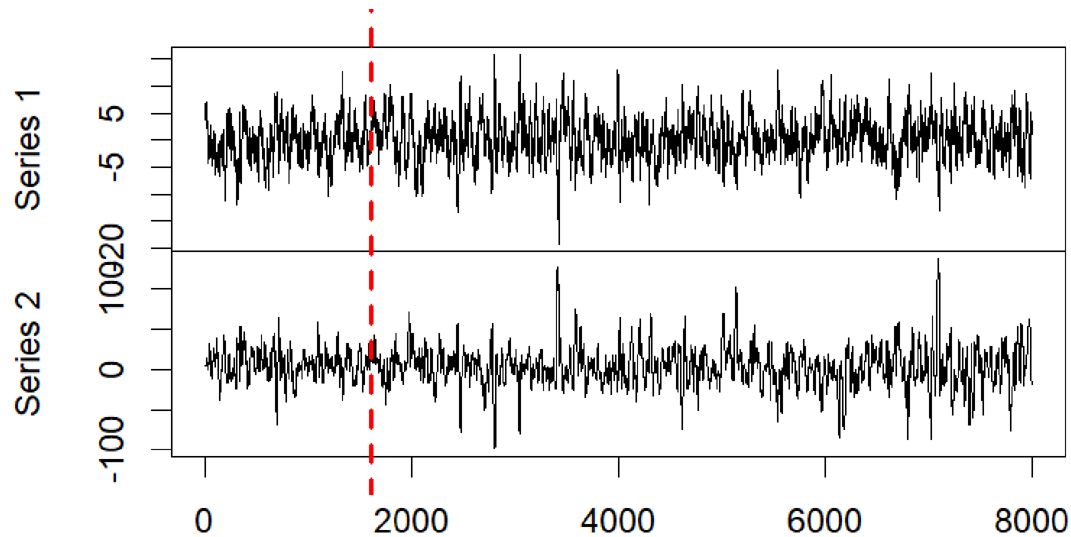

Figure 5.2: Data for the bi-variate with T = 8000 and change of mean being (0.2,0.5) at time T/5.

For change of (0.5, 0.2) with m-dependence of 20, we see that performance of the estimator, $\arg\max_{t\in[0,1]} \frac{1}{T}\tilde{S}_T(t)\hat{\Sigma}_T^{-1}\tilde{S}_T(t)$ in Theorem 4.2 decreases when the actual change-point is at $T/5$; for example, the mean absolute deviation for 8000 length data increase to 1143 from 752.7, which is for the case when the location is $T/2$. But, when the change is increased to $(0.5, 1.2)$, there is not much of a change in the performance in terms of the mean absolute deviation and in both the cases, they perform well. This is mainly because of the increase in the magnitude of change in mean; the estimation is not affected to a large extent due to the change in location in the actual change-point. For the m-dependence 10, the performance of the estimator is good for all the cases, except for the case when the location of the change-point is $T/5$ and change in mean of $(0.5, 0.2)$, but the decline of the performance is not as bad as that of the case of m-dependence 20. This is mainly due to the choice of m. Similarly, the estimation is good enough for the change in mean being $(0.5, 1.2)$ and the variation in performance is also low. We see similar observations for the datasets of length 16000, but the change in the performance is not that drastic for the m-dependence of 20 when the location of the change-point is changed from $T/2$ to $T/5$. We also present some more simulation results based on m-dependence of 30. For this case, we see that the performance of the estimator for the case when the change in mean is $(0.5, 1.2)$ is still very good, especially for dataset of length 16000. So, we can expect to obtain an increase in the accuracy of the method by allowing larger dependence based on the chosen value of T and the amount of change. Similar observations are seen in case of five-variate case when we alter the m-dependence, change in location of change-point and magnitude of change in the mean.

To show a graphical view of the results in Tables 5.1-5.4, we present Figures 5.3 and 5.4 to show the distribution of the estimated change-points once the null hypothesis is rejected. We also show the Figure 5.5 for the combinations in Table 5.1. The plots clearly demonstrate the improvement in performance when the magnitude of change is increased and m-dependence is decreased. Similar plots can also be obtained for other tables as well.

Table 5.1: Performance evaluation of the estimator on the bi-variate data, T = 8000

change	m-dependence	location	deviation	abs deviation	sqd-deviation
0.5, 0.2	10	T/2	8.3	154.1	238.5
	10	T/5	-714.7	728.9	813.02
	20	T/2	87.5	564.5	857.1
	20	T/5	-1021.1	1143	1356.9
	30	T/2	-85.33	752.7	1097
0.5, 1.2	10	T/2	-3.6	25.4	38.6
	10	T/5	-74.5	99.1	216.3
	20	T/2	45.8	144.2	214.5
	20	T/5	-226.3	280.8	358.6
	30	T/2	-95.5	352.6	551.2

Table 5.2: Performance evaluation of the estimator on the bi-variate data, T = 16000.

change	m-dependence	location	deviation	abs deviation	sqd-deviation
0.5, 0.2	10	T/2	6.1	99.6	178.2
	10	T/5	-415	522.6	687
	20	T/2	44.3	315	459.3
	20	T/5	-452	745	854
	30	T/2	96	311.2	425
0.5, 1.2	10	T/2	7	12	29
	10	T/5	-52	75	124
	20	T/2	18.3	67.8	110.2
	20	T/5	-204	213	336
	30	T/2	55	157	248

Table 5.3: Performance evaluation of the estimator on the five-variate data, T = 8000.

change	m-dependence	location	deviation	abs deviation	sqd-deviation
0.5, 0.2, 0.2, 0.5, 0.2	10	T/2	11.2	179.1	275.6
	10	T/5	-686.6	755.0	863.7
	20	T/2	93.6	573.4	848.8
	20	T/5	-913.4	1054.5	1594.8
	30	T/2	-63.4	662.7	902.0
0.5, 1.2, 0.5, 0.5, 0.5	10	T/2	-6.8	20.4	32.3
	10	T/5	-69.8	107.1	196.3
	20	T/2	53.1	137.3	250.5
	20	T/5	-241.8	312.8	375.4
	30	T/2	-88.3	322.9	560.6

Table 5.4: Performance evaluation of the estimator on the five-variate dataset, T = 16000.

change	m-dependence	location	deviation	abs deviation	sqd-deviation
0.5, 0.2, 0.2, 0.5, 0.2	10	T/2	9.7	105.8	208.1
	10	T/5	-408.2	568.7	679.2
	20	T/2	52.9	332.8	478.1
	20	T/5	-405.4	713.2	874.9
	30	T/2	121.7	351.7	471.6
0.5, 1.2, 0.5, 0.5, 0.5	10	T/2	6.7	16.2	27.3
	10	T/5	-46.8	87.3	138.7
	20	T/2	26.9	77.9	115.3
	20	T/5	-287.1	296.7	425.9
	30	T/2	61.2	146.9	255.5

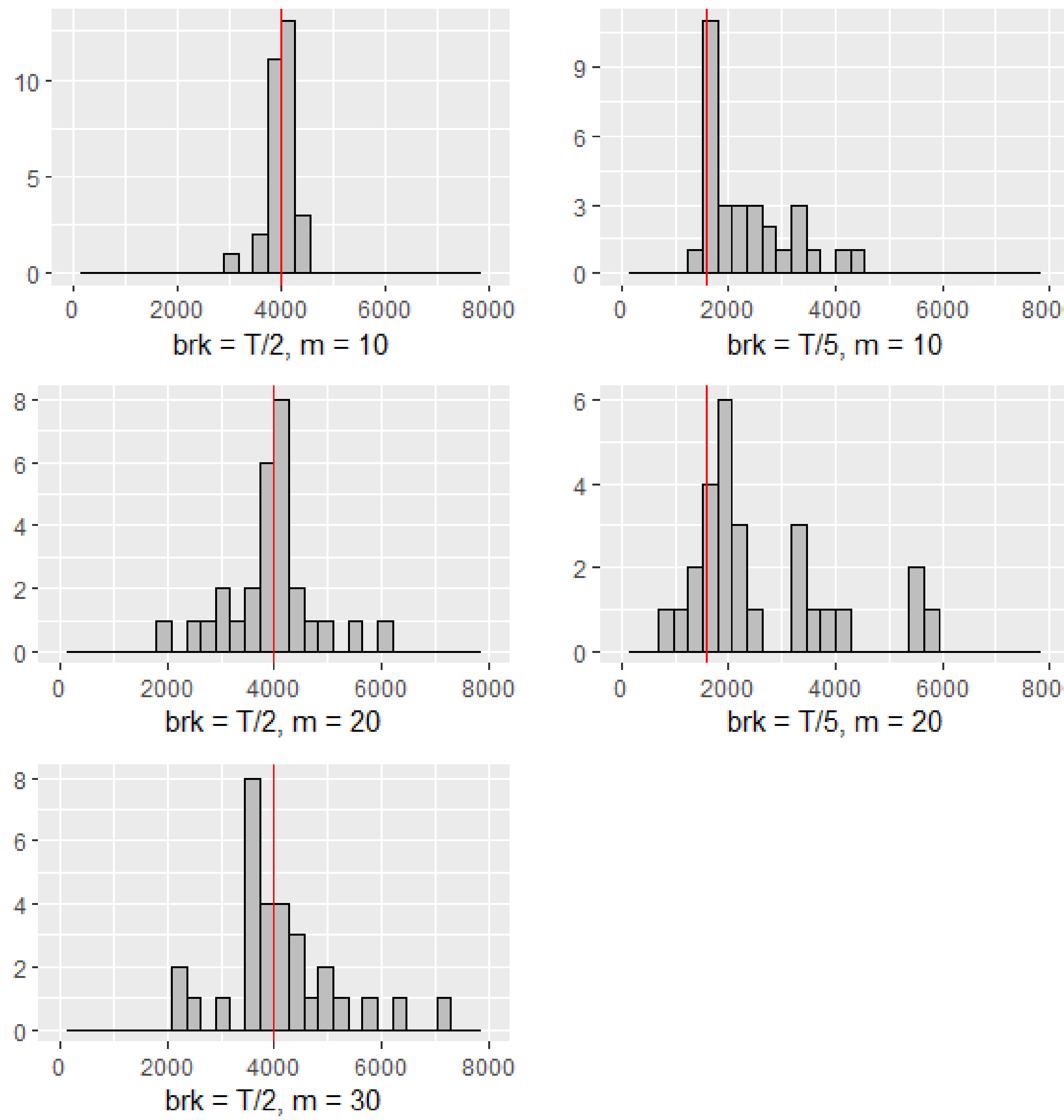

Figure 5.3: Conditional distribution of the change-point: T = 8000, Mean change = (0.2,0.5), x-axis represents the time index and y-axis represents the frequency of the estimates of the change-points.

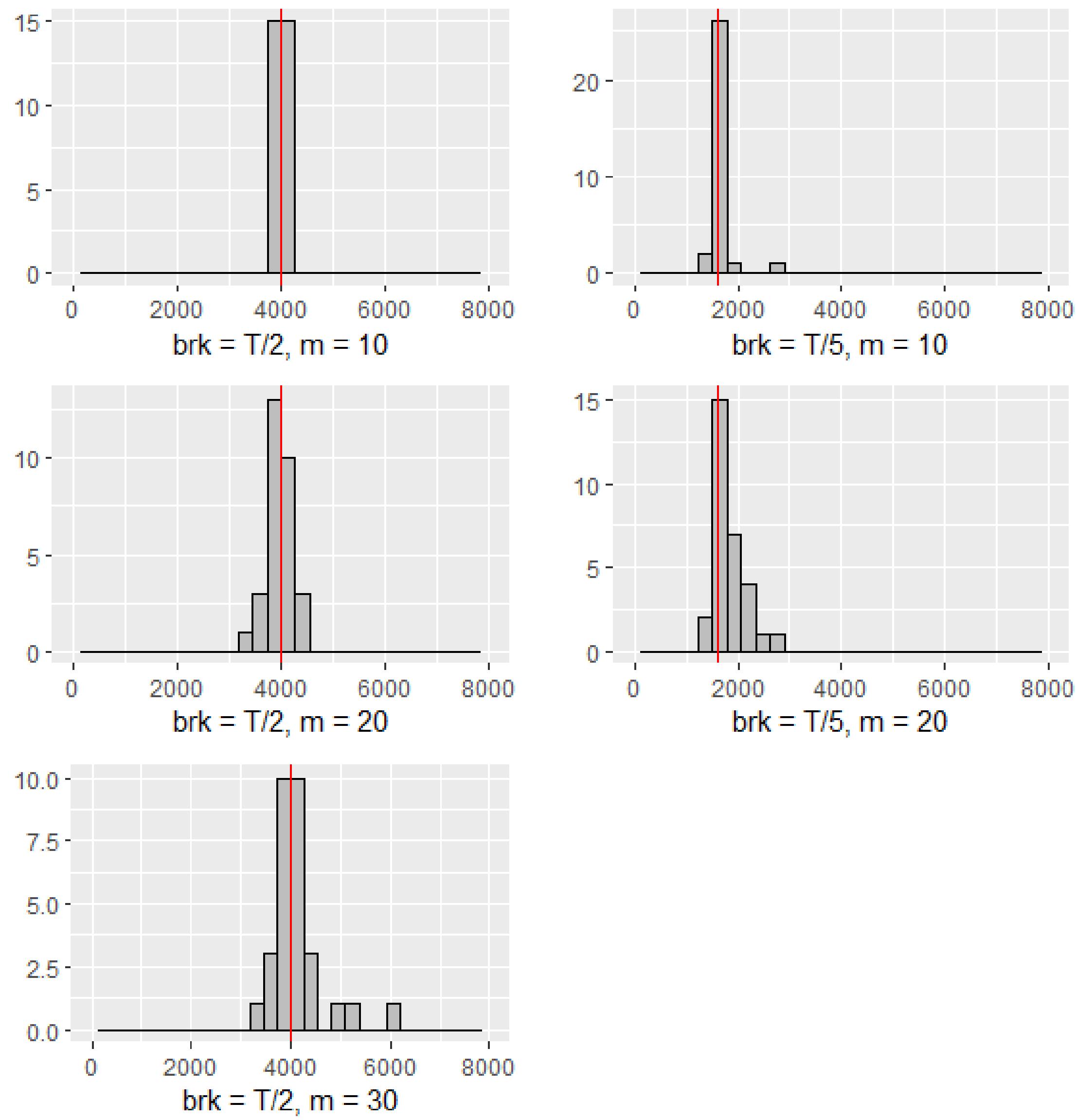

Figure 5.4: Conditional distribution of the change-point estimates: T = 8000, Mean change = (0.5,1.2), x-axis represents the time index and y-axis represents the frequency of the estimates of the change-points.

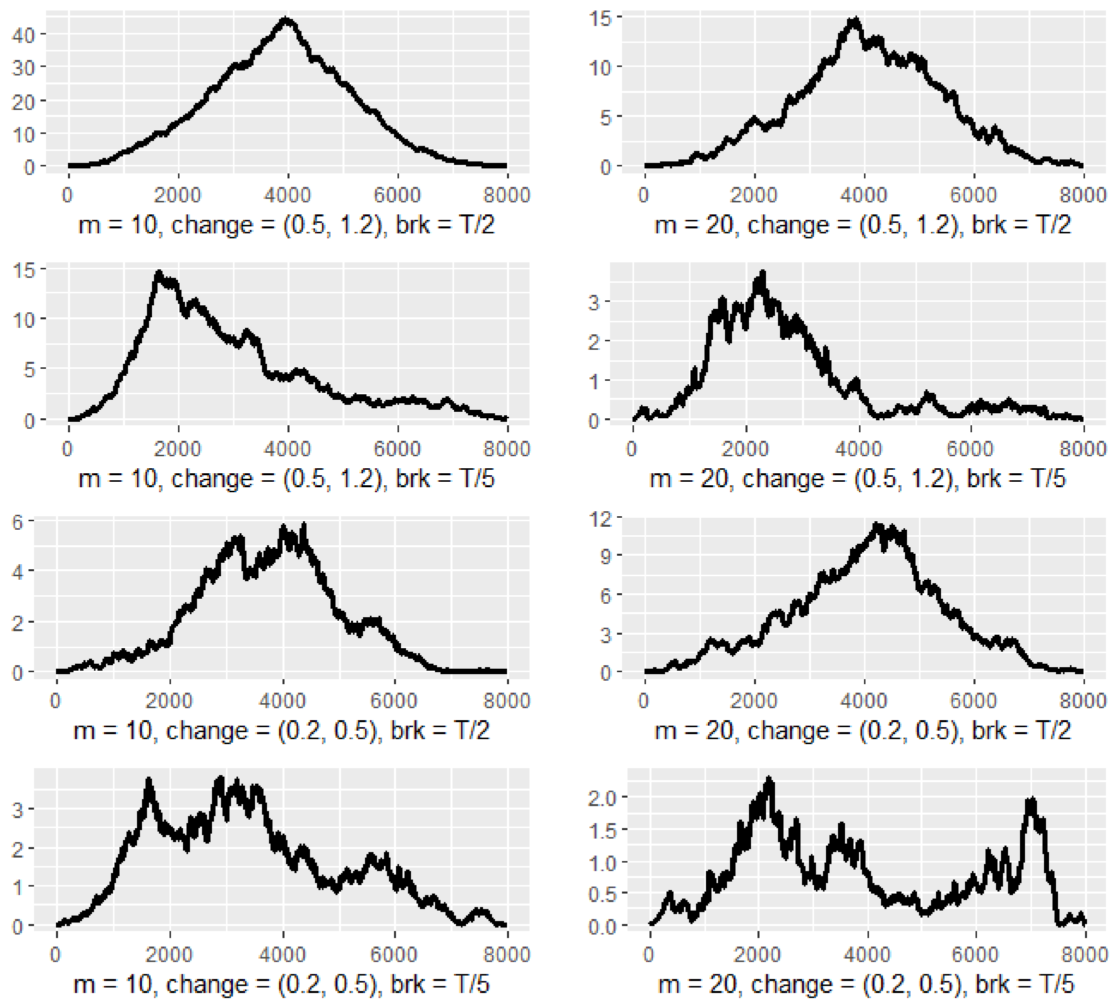

Figure 5.5: Performance of the estimator $\arg\max_{t\in[0,1]} \frac{1}{T}\tilde{S}_T(t)\hat{\Sigma}_T^{-1}\tilde{S}_T(t)$ based on one simulation for each of the cases with T = 8000.

Chapter 6

Concluding remarks

In this work, we have provided an overview of the significance and challenges associated with multivariate change-point detection methods. The review of existing methods has highlighted the different approaches employed by researchers across different disciplines, including signal processing, finance, environmental monitoring, and healthcare. From Malo's work on detecting structural breaks in multivariate time series data to Kucheva's contributions in statistical modeling, a wide array of methodologies has been discussed, each with its own strengths and limitations. After that we move to the development of the theoretical results, where we first present Theorem 4.1. The assumptions made in this theorem are mainly the assumptions that are made to perform hypothesis testing on whether there is actually a shift in the mean or not. For this, we use the statistic in Theorem 4.1, to perform the hypothesis testing. It is also worth mentioning that to be able to calculate the statistic, we found a consistent estimator $\hat{\Sigma}_N$ for Σ in Theorem 4.6. Once the testing is done and the null hypothesis gets rejected, we use Theorem 4.2 to get the consistent estimator of the change-point under certain assumptions like ergodicity, mentioned in the theorem.

These assumptions are justified by Remark 4.4, where we discuss why this assumption fits the with the model 5.0.1 with which we generate the empirical results in Chapter 5. The results points out some observations such as the effect of increase in length of the data, value of m in the m-dependence, location of the actual change-point and magnitude of the shift on the performance of the estimator. Although the method doesn't include any assumption on the dependence structure within the components, a potential limitation of the proposed method is that the Σ matrix mentioned in Theorem 4.1 may not always be positive-definite. This might pose some issue while performing the hypothesis testing as well as estimation of the change-point.

Despite the progress made in this field, there are still problems that are of interest for further exploration and development. Future research could focus on refining existing methods to handle more complex data structures, such as high-dimensional datasets or those with non-linear dependencies. An immediate question which may arise is finding the change-points based on change in distribution of innovations which in turn causes change in distribution of the data. Mattenson [11] has explored a method to detect change-points based on change in distribution of data assuming that the observations are independent, but using the linear process model with change in distribution of the innovations might be an alternative way of looking into the problem involving dependence structure of data.

Ultimately, the continued advancement of multivariate change-point detection methods holds great promise for improving our understanding of dynamic systems and in facilitating informed decision-making in a wide range of applications. Thus, this area of research provides great opportunities for researchers to push the boundaries of knowledge in this direction.

Bibliography

[1] M. Amiri, A. H. Sahari, and M. A. A. Bakar. Detecting structural breaks in time series via genetic algorithms. *International Journal of Computer Theory and Engineering*, 4(1):104–110, 2012.

[2] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.

[3] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2nd edition, 1991.

[4] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, England,, 5th edition, 2019.

[5] W. Gao, H. Yang, and L. Yang. Change points detection and parameter estimation for multivariate time series. *Soft Computing*, 24(9):6395–6407, 2020.

[6] Z. Hlávka, M. Hušková, and S. G. Meintanis. Change-point methods for multivariate time-series: paired vectorial observations. *Statistical Papers*, 61(4): 1351–1383, 2020.

[7] L. Horváth and M. Hušková. Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648, 2012.

[8] J. Kiefer. K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-V. Mises Tests. *The Annals of Mathematical Statistics*, 30(2):420 – 447, 1959.

[9] L. I. Kuncheva. Change detection in streaming multivariate data using likelihood detectors. *IEEE Transactions on Knowledge and Data Engineering*, 25(5):1175–1180, 2011.

[10] P. Malo, L. Viitasaari, O. Gorskikh, and P. Ilmonen. Non-parametric structural change detection in multivariate systems. *arXiv preprint arXiv:1805.08512*, 2018.

[11] D. S. Matteson and N. A. James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014.

[12] W. N. Meier and M. E. Mann. Climate regime shifts and multivariate analysis of ice core data. *Geophysical Research Letters*, 31(18), 2004.

[13] E. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527, 1955.

[14] C. R. S. Kotz, N. Balakrishnan and B. Vidakovic. *Encyclopedia of Statistical Sciences*, volume 6. John Wiley & Sons, 1998.

[15] W. F. Stout. *Almost Sure Convergence*. Academic Press, 1974.

[16] G. J. Szekely and M. L. Rizzo. Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of Classification*, 22(2):151–184, 2005.

[17] A. Viviani and F. Lillo. Structural breaks in time series. *Quantitative Finance*, 6(4):327–335, 2006.

[18] A. Zeileis, F. Leisch, K. Hornik, and C. Kleiber. Detecting structural breaks in multivariate time series. *Journal of Time Series Analysis*, 23(5):545–567, 2002.