STATISTICAL INFERENCE WITH HIGH-DIMENSIONAL DEPENDENT DATA

By

Shawn M. Santo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics – Doctor of Philosophy

2018

ABSTRACT

STATISTICAL INFERENCE WITH HIGH-DIMENSIONAL DEPENDENT DATA

By

Shawn M. Santo

High-dimensional time dependent data appear in practice when a large number of variables are repeatedly measured for a relatively small number of experimental units. The number of repeated measurements can range from two to hundreds depending on the application. Advances in technology have made the process of gathering and storing data such as these relatively low-cost and efficient. Demand to analyze such complex data arises in genetics, microbiology, neuroscience, finance, and meteorology. In this dissertation, we first introduce and investigate a novel solution to a classical problem that involves high-dimensional time dependent data. In addition, we propose a new approach to analyze high-dimensional dependent genomics data.

First, we consider detecting and identifying change points among covariance matrices of high-dimensional longitudinal data and high-dimensional functional data. The proposed methods are applicable under general temporospatial dependence. A new test statistic is introduced for change point detection, and its asymptotic distribution is established under two different asymptotic settings. If a change point is detected, an estimate for the location is provided. We investigate the rate of convergence for the change point estimator and study how it is impacted by dimensionality and temporospatial dependence in each asymptotic framework. Binary segmentation is applied to estimate the locations of possibly multiple change points, and the corresponding estimator is shown to be consistent under mild conditions for each asymptotic setting. Simulation studies demonstrate the empirical size and power of the proposed test and accuracy of the change point estimator. We apply our procedures on a time-course microarray data set and a task-based fMRI data set.

In the second part of this dissertation we consider a hierarchical high-dimensional dependent model in the context of genomics. Our model analyzes RNA sequencing data to identify polymorphisms with allele-specific expression that are correlated with phenotypic variation. Through simulation, we demonstrate that our model can consistently select significant predictors among a large number of possible predictors. We apply our model to an RNA sequencing and phenotypic data set derived from a sounder of swine.

ACKNOWLEDGMENTS

I would like to express the utmost gratitude to my advisor, Dr. Ping-Shou Zhong, for his assistance, support, guidance, and encouragement. Dr. Zhong taught me what it means to be a researcher in an academic environment. I would also like to thank my three committee members: Dr. Yuehua Cui, Dr. Hyokyoung Hong, and Dr. Juan Steibel. Their service is greatly appreciated. Furthermore, I would like to thank the Department of Statistics and Probability for the resources and opportunities provided to me during my Ph.D. career. Lastly, I would not have pursued a Ph.D. if it were not for Dr. M. Pádraig M. M. M^{c}Loughlin. Dr. M^{c}Loughlin challenged and encouraged me as an undergraduate student at Kutztown University of Pennsylvania in a way no professor had before.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

ASE Allele-specific expression

BOLD Blood-oxygen-level dependent

CUSUM Cumulative sum

DNA Deoxyribonucleic acid

EM Expectation-maximization

FDR False discovery rate

fMRI Functional magnetic resonance imaging

GO Gene ontology

Lasso Least absolute shrinkage and selection operator

QTL Quantitative trait loci mapping

RNA Ribonucleic acid

SNP Single-nucleotide polymorphism

CHAPTER 1

INTRODUCTION

1.1 Technology and the field of statistics

Technology is one of the chief drivers of growth and innovation in society, and its impact on the field of statistics cannot be understated. For much of the twentieth century, statisticians concentrated on solving problems in a classical setting, where the number of subjects, observations, or experimental units, exceeded the number of variables or features measured. If p is the number of variables or features, and n is the number of experimental units, then this classical setting is the so-called 'small p, large n' setting. The demand to develop robust theoretical procedures under the 'small p, large n' setting was due in large part to the data and resources available at the time. Computers were not efficient, data recording was not automated, and the scope of technology was limited; thus, there was little motivation to consider situations in which p far exceeded n. In fact, even as late as 1981, it was considered poor practice to have a study in which $n/p < 5$ (Huber 1981).

The past thirty years have been an era of accelerated technological progress in many fields in society. Biology, finance, economics, computer science, meteorology, and others, all have the available resources to gather massive amounts of information. The need to filter, understand, and analyze this information continues to grow. Data sets in numerous domain specific fields now often have more variables recorded than experimental units. This 'large p, small n' setting is what is referred to as high-dimensional data. As technology and data recording processes improve, statisticians will play an integral role in developing theoretically robust and computationally efficient statistical methods to analyze such complex data.

1.2 Low to high-dimensional data

The research in high-dimensional data has seen a shift over the past two decades from estimation to more complex forms of inference. Estimation is often an initial step in inference, but it does not allow us to quantify uncertainty. Much of the focus with regards to estimation in a high-dimensional framework has been geared towards parameter estimation in generalized linear models and graphical models (Bühlmann and van de Geer 2011). Donoho and Johnstone (1994) pioneered parameter estimation in a linear model when $p = n$. To obtain sparse estimation, Tibshirani (1996) proposed an ℓ_1-norm penalization procedure known as least absolute shrinkage and selection operator (Lasso). Under a sparsity assumption and other regularization conditions, Lasso simultaneously performs parameter estimation and variable selection. Tibshirani's seminal paper resulted in an extensive study of Lasso's theoretical properties and paved a way for valuable ℓ_1-norm and ℓ_2-norm penalization extensions. For example, Zou and Hastie (2005) introduced the elastic net to address some short-comings with regards to the number of covariates selected via Lasso. Tibshirani et al. (2005) and Yuan and Lin (2006) proposed fused Lasso and group Lasso, respectively. In 2006, Zou (2006) introduced adaptive Lasso. Fu and Knight (2000) and Zhao and Yu (2006) investigated the asymptotic behavior of Lasso-type estimates and proved under certain conditions that when the true parameter is 0, there exists non-zero probability mass at 0 for the estimator's limiting distribution. From a computation standpoint, Osborne et al. (2000) studied the primal and dual problem of Lasso, and as a result, developed a fast and efficient algorithm to obtain Lasso estimates. There is a long list of literature on regularization estimation for high-dimensional parameters. Since the main focus of this dissertation is not estimation, we do not enumerate all of them. Some important works include: Fan and Li (2001), Candes and Tao (2007), and Zhang (2010).

Inference as it relates to hypothesis testing or confidence intervals allows researchers to make scientific discoveries and improve decision making. However, statistical inference of these forms in high-dimensional data are not simple extensions of the classical inference procedures, where the number of sample subjects exceeds the number of variables measured. As was noted by Johnstone and Titterington (2009),

> *It should not, of course, be imagined that the 'large p' scenarios are mere alternative cases to be explored in the same spirit as their 'small p' forebears. A better analogy would lie in the distinction between linear and nonlinear models and methods — the unbounded variety and complexity of departures from linearity is a metaphor (and in some cases a literal model) for the scope of phenomena that can arise as the number of parameters grows without limit.*

In terms of inference for high-dimensional mean vectors, Dempster (1958) first considered a two-sample test in a $p > n$ setting. Bai and Saranadasa (1996), Chen and Qin (2010), and Cai and Xia (2014) proposed test statistics to extend the novel work of Dempster in 1958. Fujikoshi et al. (2010) provides an overview and details on testing high-dimensional mean vectors. The work on testing high-dimensional covariance matrices can be traced back to Ledoit and Wolf (2002), where they assumed p/n converges to some constant, and proved under a normality assumption that their test statistics are normal. Methodology building off Ledoit and Wolf include: Chen et al. (2010) and Cai and Ma (2013). Schott (2007), Srivastava and Yanagihara (2010), Li and Chen (2012), and Cai et al. (2013) all investigated the problem of testing the equality of high-dimensional covariance matrices for two or multiple groups. More recently, Ahmad (2017) and Zhang et al. (2018) generalized the work of Li and Chen (2012).

Some testing and confidence interval procedures with regards to Lasso estimates and generalized linear models were established by Bach (2008), Meinshausen and Bühlmann (2010), and Zhang and Zhang (2013).

To elucidate one of the challenges brought about in a high-dimensional framework, consider a classical test with regards to covariance matrices under the 'small p, large n' setting. Muirhead (2005) details a few of these tests, along with some tests for mean vectors. Suppose we are interested in testing

$$\begin{aligned} &H_0 : \Sigma_1 = \cdots = \Sigma_T \qquad \text{versus} \\ &H_1 : \text{Not all are equal,} \end{aligned} \tag{1.1}$$

where we assume X_{it} $(i = 1, \ldots, n;\ t = 1, \ldots, T)$ is a p-dimensional random vector from a multivariate normal distribution with mean μ_t and covariance Σ_t. Let, x_{it} be a realization of X_{it} from the tth population. Assume that the T populations are independent and the random sample of n vectors from each of the T populations are independent. The likelihood ratio test can be used to develop an α-level test for (1.1). The likelihood function is given by

$$L(\mu_t, \Sigma_t) = \prod_{t=1}^{T} (2\pi)^{-pn/2} |\Sigma_t|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (x_{it} - \mu_t)^{\mathrm{T}} \Sigma_t^{-1} (x_{it} - \mu_t) \right\}.$$

For observed data, $L(\mu_t, \Sigma_t)$ is a function of μ_t and Σ_t for all t. To obtain the likelihood criterion we maximize $L(\mu_t, \Sigma_t)$ under the restricted parameter space of the null hypothesis and also under the unrestricted parameter space. Let $\Omega = \{(\mu_t, \Sigma_t) : t = 1, \ldots, T\}$ and $\Omega_0 = \{(\mu_t, \Sigma_t) : \Sigma_1 = \cdots = \Sigma_T\}$ denote the unrestricted and restricted parameter spaces, respectively. Thus, the likelihood criterion is defined as

$$\lambda_n = \frac{\sup_{\Omega_0} L(\mu_t, \Sigma_t)}{\sup_{\Omega} L(\mu_t, \Sigma_t)}.$$

Let $N = Tn$, and let $A = \sum_{t=1}^{T} A_t$, where $A_t = \sum_{i=1}^{n} (x_{it} - \bar{x}_t)(x_{it} - \bar{x}_t)^{\mathrm{T}}$. For the parameter space Ω, the maximum likelihood estimators are $\hat{\mu}_{t,\Omega} = \bar{x}_t$ and $\hat{\Sigma}_{t,\Omega} = A_t/n$, for μ_t and Σ_t, respectively. For the parameter space Ω_0, the maximum likelihood estimators are $\hat{\mu}_{t,\Omega_0} = \bar{x}_t$ and $\hat{\Sigma}_{\Omega_0} = A/N$, for μ_t and Σ_t, respectively. Therefore, substituting these values back into λ_n and taking the logarithm,

$$\Lambda_n = -2\log(\lambda_n) = N \log\left(|\hat{\Sigma}_{\Omega_0}| \right) - \sum_{t=1}^{T} n \log\left(|\hat{\Sigma}_{t,\Omega}| \right).$$

Hence, an α-level test rejects H_0 of (1.1) whenever $\Lambda_n < \Lambda_\alpha$. Under an asymptotic setting when n diverges and p is fixed, the null distribution of Λ_n can be derived. Furthermore, as

$n \to \infty$, $A_t/(n-1) \to \Sigma_t$ in probability. Thus, the standard asymptotic results hold for likelihood ratios, $\Lambda_n \to \chi^2$ in distribution with degrees of freedom $(T-1)p(p+1)/2$, under the 'small p, large n' setting. However, breakdowns occur if we consider a 'large p, small n' framework.

Under a 'large p, small n' setting, Λ_n can no longer be computed and the asymptotic results are not easily extended. If $p > n$, then we can no longer compute $\log(|\hat{\Sigma}_{\Omega_0}|)$ due to A_t and A being singular. Furthermore, the asymptotic distribution under the null hypothesis is not well defined for when p diverges. In a high-dimensional framework with $p > n$, the convergence in probability of $A_t/(n-1) \to \Sigma_t$ no longer holds as demonstrated through spectral analysis by Bai and Yin (1993), Johnstone (2001), and others. As a result, testing (1.1) is not possible via a likelihood ratio test. This is just one example in which breakdowns in the classical methods occur due to an increase in data dimension. This phenomena is known as the "curse of dimensionality".

An increase in data dimension can produce extra noise, computation challenges, and a failure in many of the existing classical statistical procedures. However, in certain situations an increase in dimensionality may be a blessing (Donoho 2000). For further challenges associated with high-dimensional data we encourage readers to see Fan and Li (2006) and Fan et al. (2014a).

1.3 Independent to dependent data

The likelihood ratio test for (1.1) as described in Section 1.2 further breaks down if the T groups are not independent. In this dissertation, measurements of a sample that are repeatedly recorded will be referred to as longitudinal data when the number of repeated measurements is small. If the number of repeated measurements is large, or dense, we will refer to it as functional data. Measurements taken over time allow researchers to understand the evolution of the sample subjects, detect and identify changes in certain variables across time, and study sequences of events. In longitudinal or functional data sets, temporal dependence exists among measurements from the same subject, and adds a layer of complexity to the theoretical and computational analysis. Methodology developed under a T-independent sample framework is not applicable for a T-dependent sample. For example, Chen and Qin (2010) and Li and Chen (2012) considered an independent two-sample high-dimensional test for mean vectors and covariance matrices, respectively. However, their methods are not applicable in a temporal dependent setting. There are two types of dependence in the data: temporal and spatial. If these dependencies are ignored, then inference procedures are invalid and misleading. Currently, there is no existing work accounting for the aforementioned dependencies in high-dimensional covariance testing and change point detection and identification. The asymptotic analysis is more complicated when both dependencies are considered. Generalizing to an asymptotic framework for high-dimensional functional data further increases complexity.

1.4 Change point detection and identification

Given (1.1) for time dependent data, two questions naturally arise. First: Can we detect changes among T dependent covariance matrices? Second: Can we identify the time points for where those changes occur? These questions have profound effects for time dependent data. Their answers can provide critical information to individuals in the fields of finance, genetics, neuroscience, climatology, and more.

Change point detection is a classical problem in time series analysis. Numerous supervised and unsupervised machine learning algorithms are used in various change point detection applications. Aminikhanghahi and Cook (2016) detail a few multi-class supervised learning algorithms such as Gaussian mixture models, hidden Markov models, and decision trees. Their work also highlights likelihood ratios, probabilistic models, graphs, and clustering as further approaches to the change point detection problem. One of the most common techniques in change point detection is the cumulative sum (CUSUM) method by Page (1954). Measurements in a process are cumulatively summed according to a weighted procedure. A change point is identified once the cumulative sum quantity exceeds a threshold value. Chernoff and Zacks (1964) laid the groundwork for change point detection with regards to the mean of normal random variables. Accordingly, a series of methodologies were developed in independent univariate and multivariate settings. Some of these works include: Kander and Zacks (1966), Yao and Davis (1986), Sen and Srivastava (1973), and Srivastava and Worsley (1986). Chapter 2 in both Csörgö and Horváth (1997) and Brodsky and Darkhovsky (1993) detail nonparametric change point detection methods based on Wilcoxon-type statistics, U-type statistics, and M-estimators. Johnson and Bagshaw (1974), Brown et al. (1975), and Horváth and Kokoszka (1997) introduced methods to address the change point problem for dependent data. For further details on classical change point detection and identification procedures, we refer readers to Basseville and Nikiforov (1993) and Brodsky (2017).

In terms of a classical procedure for testing (1.1) with T dependent groups, there is none. A multivariate procedure to test (1.1) was proposed by Aue et al. (2009). Assume X_t $(t = 1, \ldots, T)$ are p-dimensional temporal dependent random vectors from a multivariate distribution with mean μ and covariance Σ_t. Thus, x_t is an observation at the tth time point. To test (1.1), Aue et al. (2009) considered the quantity, S_k $(k = 1, \ldots, T)$ such that

$$S_k = \frac{1}{\sqrt{T}} \left\{ \sum_{j=1}^{k} \mathrm{vech}(x_j x_j^{\mathrm{T}}) - \frac{k}{T} \sum_{j=1}^{T} \mathrm{vech}(x_j x_j^{\mathrm{T}}) \right\},$$

where for any $p \times p$ symmetric matrix M, $\mathrm{vech}(M)$ represents the stacked columns of the lower triangular region of M in the form of a $p(p-1)/2$ vector. The quantity S_k was motivated by the fact that under H_0 of (1.1), $E\{\mathrm{vech}(x_j x_j^{\mathrm{T}})\} = E\{\mathrm{vech}(x_i x_i^{\mathrm{T}})\}$ for all $i, j \in \{1, \ldots, T\}$. Based on S_k, they introduced a test statistic $\Omega_T = T^{-1} \sum_{k=1}^{T} S_k^{\mathrm{T}} \hat{\Sigma}_T^{-1} S_k$, where $\hat{\Sigma}_T$ is an estimator such that $|\hat{\Sigma}_T - \Sigma_T|_E = o_p(1)$ as T diverges, and for any matrix A, $|A|_E = \sup_{x \neq 0} |Ax|/|x|$. They derived the test statistic's asymptotic distribution under the null hypothesis with $T \gg p$.

However, Aue et al. (2009)'s method fails in a high-dimensional framework since $\hat{\Sigma}_T$ is not invertible if $p \gg T$. In addition, Aue et al. did not consider a setting in which $n > 1$, and thus, their methodology does not permit multiple-subject inference. In Section 1.2 we highlighted the fact that recent research has addressed the high-dimensional challenges for testing (1.1) but not dependence. In this section we detailed a procedure that incorporates dependence but not a 'large p, small n' framework. Therefore, a gap exists. How can we test (1.1) for high-dimensional time dependent data?

1.5 High-dimensional time dependent data

High-dimensional longitudinal data appear in practice when a large number of variables, p, are repeatedly measured for a relatively small number of experimental units, n. The number of repeated measurements, T, can range from two to hundreds depending on the application. Throughout this dissertation, longitudinal data will refer to settings when T is small. High-dimensional functional data will refer to settings in which T is large or dense. For details on functional data analysis we refer readers to Ramsay and Silverman (2005).

Consider an experiment where patients have their gene expressions measured throughout the course of a treatment regimen. Doctors and clinicians may be interested in understanding how these gene expressions are regulated over time. In studies such as this, the number of gene expressions, p, measured is anywhere from a few hundred to a few thousand, and the number of patients, n, along with the number of repeated measurements, T, is small. We will refer to this as high-dimensional longitudinal data. As another example, consider a functional magnetic resonance imaging (fMRI) study where patients have their brain activity measured while performing various tasks. Thousands of blood-oxygen-level dependent (BOLD) responses are recorded, hundreds of times during the duration of a scan, for voxels corresponding to regions of interest in the patient's brain. For this single patient, radiologists may be interested in identifying and understanding significant spatial and temporal changes. The BOLD data from an fMRI experiment are considered high-dimensional functional data.

1.6 Dissertation outline

In Chapters 2 and 3 of this dissertation we develop and evaluate a procedure to test (1.1) for high-dimensional longitudinal and high-dimensional functional data, respectively. To visualize our objective in a high-dimensional longitudinal setting, consider Figure 1.1. Each sub-plot represents the covariance matrix at the respective time point. From Figure 1.1 it is clear that the covariance is homogeneous between time points one through three; there is a different covariance structure at $t = 4$; for time points five and six the covariance structure is homogeneous again.

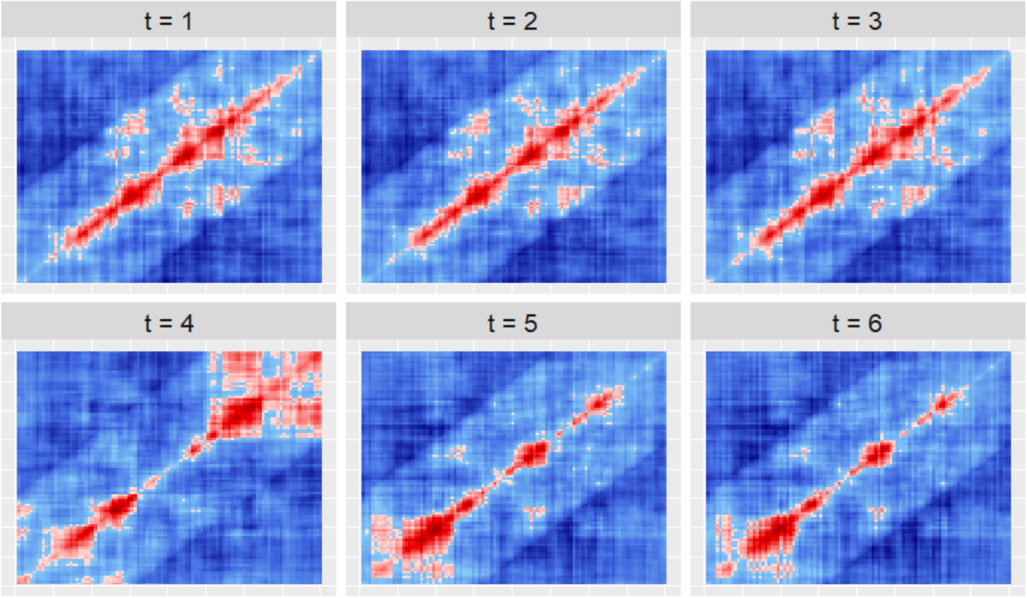

Figure 1.1: Population covariance heat maps at six time points. Change points exist at time $t = 3$ and at time $t = 4$.

Our statistical test will first detect the presence of any change points among the T covariance matrices. If we can conclude that change points exist, we further identify the time points at which changes occur. The procedures we propose are pioneering with regards to (1.1) for high-dimensional longitudinal and high-dimensional functional data. As is discussed in detail in Chapters 2 and 3, some research has provided a solution to test (1.1) in a high-dimensional framework, but no method has been developed for high-dimensional time dependent data. In addition to the theoretical challenges, we also address the natural computation challenges that arise with such massive time dependent data. We ensure our method is practical and accessible to the end users in biology, neuroscience, and other fields via an R package.

In Chapter 4 we consider a different type of high-dimensional dependent data, where we propose a novel hierarchical model for genomics applications. Our interest is to link a phenotypic response with single nucleotide polymorphisms (SNPs) that have allele-specific expression (ASE). To account for dependence among the latent genotype and ASE status combination, we consider a hidden Markov model and incorporate regularized regression to address the high-dimensionality. Our problem can be depicted with the graphical model in Figure 1.2 for the ith individual with five SNPs. Let X_{il}, G_{il}, δ_{il} be the RNA read counts, genotype and ASE status, and allele-specific expression ratio, respectively for the ith individual at the lth SNP. Let Y_i be an observed phenotypic response. Given the relationships between X, G, and δ we first aim to estimate the latent variables G_{il} and δ_{il} given X and an assumed Markov structure for G. For an observed phenotypic response, Y, we use regularized regression to select the significant δs.

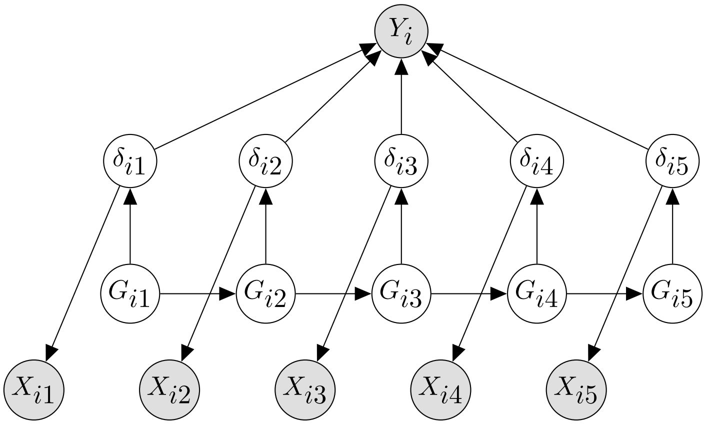

Figure 1.2: A small graphical model for the problem considered in Chapter 4. Grey circles represent observed values. White circles represent latent variables.

In, Chapter 5, we discuss possible theoretical and computational extensions to the results of Chapters 2 – 4.

All proofs to lemmas and theorems are provided in the sections titled “Technical details” of the respective chapter.

CHAPTER 2

HOMOGENEITY TESTS OF COVARIANCE MATRICES WITH HIGH-DIMENSIONAL LONGITUDINAL DATA

2.1 Introduction

In a typical time-course microarray data set, thousands of gene expression values are measured repeatedly from the same subject at different stages in a developmental process (Tai and Speed, 2006). As a motivating example, Taylor et al. (2007) conducted a longitudinal study on 69 patients infected with hepatitis C virus. Their gene expression values were measured once before treatment and five times during the treatment regimen of pegylated alpha interferon and ribavirin. One purpose of the study was to identify which genes were regulated by treatment. The repeated measurements enable researchers to understand gene regulation over time. An important task in genomic studies is to identify gene sets with significant temporal changes (Storey et al., 2005). Much evidence has shown that gene interaction and co-regulation play a critical role in the etiology of various diseases (Shedden and Taylor, 2005). One application of our methods is to identify gene sets with significant changes in their covariance matrices, because the covariance matrix or its inverse can be used for quantifying interaction and co-regulation among genes (Danaher et al., 2015).

Assume that $Y_{it} = (Y_{it1}, \dots, Y_{itp})^{\mathrm{T}}$ is a p-dimensional random vector with mean μ_t and covariance Σ_t. In the aforementioned applications, Y_{it} ($i = 1, \dots, n$; $t = 1, \dots, T$) represents gene expressions for p genes in a gene set measured from the ith individual at the tth developmental stage, where n is the sample size and T is the total number of finite stages. The number of genes, p, in a given gene set ranges from a hundred to a few thousand, as illustrated by the histogram in Figure 2.2 in Section 2.6, but n and T are small in the study. Thus, p can be much larger than n and T. We focus on testing the homogeneity of covariance matrices:

$$H_0 : \Sigma_1 = \cdots = \Sigma_T \quad \text{versus} \quad H_1 : \Sigma_k \neq \Sigma_l \tag{2.1}$$

for some $1 \leq k \neq l \leq T$. The alternative in (2.1) can be written as a change point type alternative:

$$H_1 : \Sigma_1 = \cdots = \Sigma_{k_1} \neq \Sigma_{k_1+1} = \cdots = \Sigma_{k_q} \neq \Sigma_{k_q+1} = \cdots = \Sigma_T, \tag{2.2}$$

where $1 \leq k_1 < \cdots < k_q < T$ are unknown locations of change points. This alternative is of interest in practice because it specifies the locations of changes. For example, researchers are often interested in understanding dynamic gene regulation. By identifying the change points, we can infer the change pattern of gene regulation, which is important for developing diagnostic and preventive tools for some diseases (Koh et al., 2014).

Testing the homogeneity of covariance matrices is a classical problem in multivariate analysis. Classical methods for testing (2.1) include the likelihood ratio test (Muirhead, 2005) and Box's M test (Box, 1949). Some resampling methods have also been proposed by Zhang and Boos (1992) and Zhu et al. (2002). However, these methods are not valid for the aforementioned applications for the following reasons. First, these methods require n to be much larger than p. Thus, they are not applicable under the large p, small n paradigm. Second, these methods are only valid for independent samples without temporal dependence, but the independence assumption is not valid for high-dimensional longitudinal data because the repeated measurements obtained from the same individual are temporally dependent.

There is some existing research on testing (2.1) in the large p, small n scenario for independent samples. Li and Chen (2012) considered testing the equality of two covariance matrices for two independent samples. Schott (2007) and Srivastava & Yanagihara (2010) proposed test statistics for (2.1) based on estimators of the summation of the weighted pairwise Frobenius norm distances between any two covariance matrices. Zheng et al. (2015) and Yang and Pan (2017) applied random matrix theory to test the equality of two large-dimensional covariance matrices.

Some methods have also been developed in neuroscience literature under the large p and large T setup with $T > p$, which is different from our large p, small n and T setup. For example, Barnett & Onnela (2016) proposed a sieve bootstrap covariance change point detection method that requires removing both boundaries of a time series with length greater than p to avoid ill-conditioned covariance matrices. Laumann et al. (2017) discussed a method for detecting changes in covariances by assessing the stability of multivariate kurtosis using a simulation approach. Their methods also require $T > p$ to ensure the existence of an inverse of a sample covariance matrix. In addition to the aforementioned multivariate detection procedures, a marginal pair-wise testing procedure was developed by Zalesky et al. (2014). Their approach relies on a sliding window to detect changes in correlation coefficients between a pair of coordinates. The p-value for each pair is obtained by resampling residuals after fitting vector autoregressive models. It is then applied to test the homogeneity of covariance matrices using multiple testing. Despite the above advances, no existing multivariate method can be applied directly to test (2.1) for temporal dependent data under the large p, small n and T setup.

This chapter proposes a new method for testing the equality of covariance matrices with high-dimensional longitudinal data under the large p, small n and T scenario. The proposed method considers both spatial and temporal dependence. Spatial dependence refers to the dependence among different components of Y_{it}, and temporal dependence refers to the dependence between Y_{it} and Y_{is} for any two time points $t \neq s$. The asymptotic distribution of the proposed test statistic is derived under mild conditions on dependence without any explicit requirement on the relationships between p, n and T.

We also propose a method for estimating the location of change points $k_1, \ldots, k_q$ among covariance matrices. There exists some work on identifying change points in high-dimensional means, but the literature for high-dimensional covariances is very small. Aue et al. (2009) laid groundwork by considering a p-dimensional multivariate, possibly high-dimensional, time series setup where T diverges, $n = 1$ and $p < T$. Their test statistic involves the inverse of a $p \times p$ sample covariance matrix, which is singular if $p > T$. Thus, their method is not applicable to high-dimensional longitudinal data. In the case with finite p and n but diverging T, one major concern is that the change point estimator is not consistent (Hinkley, 1970) and only the ratios k_i/T $(i = 1, \ldots, q)$ are consistent. When p is finite but $n \to \infty$, it has been shown that change points can be estimated consistently. However, it is not clear how the data dimension affects the rate of convergence. We study the rate of convergence of our proposed change point estimator and find that it depends on the data dimension, sample size, noise level and signal strength. Consistency of the change point estimator is possible even in the high-dimensional case. Furthermore, we propose a binary segmentation procedure for identifying the locations of multiple change points, whose consistency is also established.

Our work is related to, but different from, that of Li and Chen (2012), who considered a test for the equality of two covariance matrices with two independent samples. First, we consider a general homogeneity test of covariance matrices with more than two populations, while Li and Chen only considered a two-sample case. Second, Li and Chen considered the test for two independent samples, but our proposal can accommodate both temporal and spatial dependence. Moreover, our method is designed to test for the existence of change points among high-dimensional covariance matrices for longitudinal data. Therefore, the test procedure considered in this chapter is different from that in Li and Chen (2012).

This chapter makes the following contributions. From a methodology perspective, the proposed test procedure provides a novel solution for change point detection problems in the large p, small n and T scenario. The test statistic combines the strength of maximal and Frobenius norms, and is powerful against the alternative. Second, we propose a method for estimating locations of change points among high-dimensional covariance matrices. The proposed change point detection and identification procedures are widely applicable without any sparsity assumption. We establish the asymptotic distribution of a test statistic for data with general temporal and spatial dependence. The identification procedure for multiple change points is shown to be consistent. Our results reveal the impact of data dimension, sample size, and signal-to-noise ratio on the rate of convergence of the change point estimator. The proposed methods formally address two challenges that are unsolved in the existing covariance change point literature: the large p, small n and T issue, and spatial and temporal dependence.

The remaining sections of this chapter are organized as follows. Section 2.2 details our basic settings with regards to covariance testing. In Section 2.3 we introduce our testing procedure and test statistics along with their asymptotic distributions. Section 2.4 introduces an estimator for change point identification. Moreover, binary segmentation is proposed to identify multiple change points. Sections 2.5 and 2.6 demonstrate the finite sample performance of our procedures via simulation and analysis of a time-course microarray data set, respectively. All proofs of theorems and necessary lemmas are available in Section 2.7.

2.2 Basic setting

Let $Y_{it} = (Y_{it1}, \ldots, Y_{itp})^{\mathrm{T}}$ be the observed p-dimensional random vector for the ith individual at time point $t = 1, \ldots, T$, where $T \geq 2$, and $i = 1, \ldots, n$. Assume that Y_{it} follows the model

$$Y_{it} = \mu_t + \varepsilon_{it}, \tag{2.3}$$

where μ_t is a p-dimensional unknown mean vector and $\varepsilon_{it} = (\varepsilon_{it1}, \ldots, \varepsilon_{itp})^{\mathrm{T}}$ is a multivariate normally distributed random error vector with mean zero and covariance $\mathrm{var}(\varepsilon_{it}) = \Sigma_t$. A generalization to the non-Gaussian setup is given in Section 2.5. In addition, it is assumed that $\varepsilon_{it} = \Gamma_t Z_i$ for a $p \times m$ matrix Γ_t, where $m \geq pT$, and Z_i is an m-dimensional standard multivariate normally distributed random vector so that $\mathrm{cov}(\varepsilon_{is}, \varepsilon_{jt}) = \Gamma_s \Gamma_t^{\mathrm{T}} = C_{st}$ if $i = j \in \{1, \ldots, n\}$ and is 0 if $i \neq j$. The random errors $\{\varepsilon_{it}\}_{i=1}^{n}$ are independent, but $\{\varepsilon_{it}\}_{t=1}^{T}$ depend on each other. Of interest is to test whether any change points among covariances occur at some time points $t \in \{1, \ldots, T-1\}$. We test the hypothesis H_0 versus H_1 specified in (2.1) and (2.2). If H_0 is rejected, we further estimate the locations of change points.

2.3 Homogeneity tests of covariance matrices

At each $t \in \{1, \ldots, T-1\}$, we define a measure $D_t = w^{-1}(t) \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2})^2\}$, where $w(t) = t(T-t)$. Measure D_t characterizes the differences among the covariances before t and after t. Clearly, $D_t = 0$ for all $t \in \{1, \ldots, T-1\}$ under H_0, and $D_t \neq 0$ for any t under H_1. Therefore, $\max_{1 \le t \le T-1} D_t = 0$ under H_0, and $\max_{1 \le t \le T-1} D_t > 0$ under H_1. Thus, D_t is useful for distinguishing the null and alternative hypotheses.

Measure D_t is different from measure $S_{1,T} = \sum_{s_1=1}^{T-1} \sum_{s_2=s_1+1}^{T} \mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2})^2\}$ used in Schott (2007), who applied $S_{1,T}$ in constructing a homogeneity test specified in (2.1) for independent samples. In fact, for any $t \in \{1, \ldots, T-1\}$, $D_t = S_{1,T} - (S_{1,t} + S_{t+1,T})$, where $S_{1,t}$ and $S_{t+1,T}$ quantify the differences among covariances only before time t and only after time t, respectively. These are not useful for measuring the differences among covariances before and after time t. Measure D_t removes both $S_{1,t}$ and $S_{t+1,T}$ from $S_{1,T}$.

To construct an unbiased estimator of D_t, we need an unbiased estimator of $\mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})$. We make use of U-statistic type estimators because they avoid bias that is not ignorable in a high- dimensional setup (Bai & Saranadasa, 1996; Chen & Qin, 2010). Otherwise, bias correction could be a challenge and require conditions on the data dimension and sample size that limit the scope of applications. Let $\tilde{\sum}$ denote summation over mutually different indices of sample subjects. For example, $\tilde{\sum}_{i,j,k}$ means summation over $\{(i,j,k) \in \{1, \ldots, n\} : i \neq j, j \neq k, k \neq i\}$. For any $s_1, s_2 \in \{1, \ldots, T\}$, define $U_{s_1 s_2,0} = (1/P_n^2) \sum_{i \neq j}^{n} (Y_{is_1}^{\mathrm{T}} Y_{js_2})^2$ as an unbiased estimator of $\mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2}) + \mu_{s_1}^{\mathrm{T}} \Sigma_{s_2} \mu_{s_1} + \mu_{s_2}^{\mathrm{T}} \Sigma_{s_1} \mu_{s_2} + (\mu_{s_1}^{\mathrm{T}} \mu_{s_2})^2$ where $P_n^k = n!/(n-k)!$. To remove the nuisance terms $\mu_{s_1}^{\mathrm{T}} \Sigma_{s_2} \mu_{s_1}$ and $(\mu_{s_1}^{\mathrm{T}} \mu_{s_2})^2$, we define $U_{s_1 s_2,1} = (1/P_n^3) \tilde{\sum}_{i,j,k} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1}$ as an unbiased estimator of $\mu_{s_1}^{\mathrm{T}} \Sigma_{s_2} \mu_{s_1} + (\mu_{s_1}^{\mathrm{T}} \mu_{s_2})^2$ and, similarly, $U_{s_2 s_1,1}$ is an unbiased estimator of $\mu_{s_2}^{\mathrm{T}} \Sigma_{s_1} \mu_{s_2} + (\mu_{s_1}^{\mathrm{T}} \mu_{s_2})^2$. To remove the nuisance term $(\mu_{s_1}^{\mathrm{T}} \mu_{s_2})^2$, we define $U_{s_1 s_2,2} = (1/P_n^4) \tilde{\sum}_{i,j,k,l} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{ks_1}^{\mathrm{T}} Y_{ls_2}$ as an unbiased estimator of $(\mu_{s_1}^{\mathrm{T}} \mu_{s_2})^2$. A computation efficient formulation of $U_{s_1 s_2,1}$ and $U_{s_1 s_2,2}$ is given in the Appendix. Finally, we define an unbiased estimator for $\mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})$ as

$$U_{s_1 s_2} = U_{s_1 s_2,0} - U_{s_1 s_2,1} - U_{s_2 s_1,1} + U_{s_1 s_2,2}. \tag{2.4}$$

The estimator $U_{s_1 s_2}$ is a generalization of the estimator for the trace of the covariance given by Chen et al. (2010) and Li and Chen (2012). For $t = 1, \ldots, T-1$, an unbiased estimator of D_t is

$$\hat{D}_{nt} = \frac{1}{w(t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} (U_{s_1 s_1} + U_{s_2 s_2} - U_{s_1 s_2} - U_{s_2 s_1}). \tag{2.5}$$

To study the asymptotic variance of $\hat{D}_{nt}$ for $t = 1, \ldots, T-1$, define

$$V_{0t} = \sum_{\substack{s_1, s_2, \\ h_1, h_2}}^{*} \sum_{\substack{u, v, \\ k, l \in \{1,2\}}} (-1)^{|u-v|+|k-l|} \mathrm{tr}^2(C_{s_u h_k} C_{s_v h_l}^{\mathrm{T}})$$

and

$$V_{1t} = \sum_{\substack{s_1, s_2, \\ h_1, h_2}}^{*} \sum_{u, k \in \{1,2\}} (-1)^{|u-k|} \mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2}) C_{s_u h_k} (\Sigma_{h_1} - \Sigma_{h_2}) C_{s_u h_k}^{\mathrm{T}}\},$$

where $\sum^{*}_{\substack{s_1, s_2, \\ h_1, h_2}} = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{h_1=1}^{t} \sum_{h_2=t+1}^{T}$. If no temporal dependence exists, then $C_{s_u h_k} = 0$ for any $s_u \neq h_k$, and $V_{0t} = \sum^{*}_{s_1, s_2} \sum_{u,v \in \{1,2\}} \mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v})$ where $\sum^{*}_{s_1, s_2} = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T}$. Up to a scale factor, this V_{0t} is the part of the variance of $\hat{D}_{nt}$ for the case with independent samples under H_0 .

The asymptotic setting considered in this chapter is $p(n) \to \infty$ as $n \to \infty$, where p is considered to be a function of n. We do not require a specific relationship between p and n. Instead, for any $t \in \{1, \ldots, T-1\}$, we have two regularity conditions. For any matrix A, denote $A^{\otimes 2} = AA^{\mathrm{T}}$. Then:

Condition 1. $\mathrm{tr}\{(\Gamma_{s_2}^{\mathrm{T}} C_{s_1 h_1} \Gamma_{h_2})^{\otimes 2}\} = o(V_{0t})$ *for any* $s_1, s_2, h_1, h_2 \in \{1, \ldots, T\}$*;*

Condition 2. $\mathrm{tr}\big[\{(\Gamma_{s_1} + \Gamma_{s_2})^{\mathrm{T}} (\Sigma_{s_1} - \Sigma_{s_2})(\Gamma_{s_1} - \Gamma_{s_2})\}^{\otimes 2}\big] = o(nV_{1t})$ *for* $s_1 \in \{1, \ldots, t\}$ *and* $s_2 \in \{t+1, \ldots, T\}$.

Condition 1 generalizes Condition 2 imposed by Li and Chen (2012) to a T-sample test with temporal dependence. If there is no temporal dependence, Condition 1 can be simplified to $\mathrm{tr}(\Sigma_{s_2}\Sigma_{s_1}\Sigma_{h_2}\Sigma_{s_1}) = o(V_{0t})$. In general, the left-hand side of the equality in Condition 1 is bounded by $\{\mathrm{tr}(\Sigma_{h_2}\Sigma_{h_1}\Sigma_{h_2}\Sigma_{h_1})\mathrm{tr}(\Sigma_{s_2}\Sigma_{s_1}\Sigma_{s_2}\Sigma_{s_1})\}^{1/2}$, which is of order $O(p)$ if all the eigenvalues of Σ_t are bounded. If the temporal dependence is not overwhelming so that $V_{0t} \asymp p^{\delta}$ for any $\delta > 1$, then Condition 1 holds. To appreciate this point, consider a null hypothesis case with $C_{st} = (1 - r_{st,n})\Sigma$ for $s, t \in \{1, \ldots, T\}$. Here $1 - r_{st,n}$ measures the temporal correlation. If $r_{st,n}$ is small for all s, t, then the temporal dependence among $\{Y_{it}\}_{t=1}^{T}$ is strong. Let $r_n = \sum^{*}_{s_1,s_2,h_1,h_2} \sum_{u,v,k,l\in\{1,2\}} (-1)^{|u-v|+|k-l|} r_{s_u h_k,n} r_{s_v h_l,n}$. If $r_{st,n} \to 0$ for all s, t, then $V_{0t} \asymp r_n \mathrm{tr}^2(\Sigma^2) \asymp r_n p^2$ provided all the eigenvalues of Σ are bounded. If the temporal dependence is not too strong so that $1/p = o(r_n)$, then Condition 1 holds as $p \to \infty$. Intuitively, Condition 1 implies that spatial and temporal dependence cannot be too strong.

Condition 2 is automatically true under H_0 because its left-hand side equals zero. Hence, it is not needed under H_0. If there is no temporal dependence, it can be shown that the left-hand side of Condition 2 is $\mathrm{tr}\big\{(\Sigma_{s_1}^2 - \Sigma_{s_2}^2)^2\big\}$, whose order is not larger than V_{1t}. Therefore, Condition 2 is not needed for data without temporal dependence. This condition implies that the alternatives should not be too far away from the null hypothesis. Otherwise, the alternatives are easy to detect because the test statistics would diverge to infinity.

Theorem 1 states the mean and variance of $\hat{D}_{nt}$. The proof is given in Section 2.7.

Theorem 1. *The expectation of* $\hat{D}_{nt}$ *is* $E(\hat{D}_{nt}) = D_t$. *Under Condition 1, the leading order variance of* $\hat{D}_{nt}$ *is* $\sigma_{nt}^2 = w^{-2}(t)\big(4V_{0t}/n^2 + 8V_{1t}/n\big)$.

Based on Theorem 1, we observe that $E(\hat{D}_{nt}) = D_t = 0$ under H_0. Under alternative H_1 in (2.2), it is clear that $E(\hat{D}_{nt}) > 0$ for all t under H_1. Therefore, $\hat{D}_{nt}$ is able to distinguish the null and alternative hypotheses in (2.1) and (2.2).

If $T = 2$ and no temporal dependence exists, V_{0t} and V_{1t} are, respectively, simplified to $V_{01} = \mathrm{tr}^2(\Sigma_1^2) + 2\mathrm{tr}^2(\Sigma_1\Sigma_2) + \mathrm{tr}^2(\Sigma_2^2)$ and $V_{11} = \sum^{*}_{s_1,s_2}\sum_{u=1}^{2} \mathrm{tr}\big[\{\Sigma_{s_u}(\Sigma_{s_1} - \Sigma_{s_2})\}^2\big]$, which are the same as those obtained by Li and Chen (2012). For a general case with temporal dependence, $V_{01} = \mathrm{tr}^2(\Sigma_1^2) + 2\mathrm{tr}^2(\Sigma_1\Sigma_2) + \mathrm{tr}^2(\Sigma_2^2) - 4\{\mathrm{tr}^2(\Sigma_1 C_{21}) + \mathrm{tr}^2(\Sigma_2 C_{12})\} +$

$2\{\mathrm{tr}^2(C_{12}C_{12}^{\mathrm{T}}) + \mathrm{tr}^2(C_{12}C_{12})\}$. The last four terms in V_{01}, due to the temporal dependence, are not included in Li and Chen's test. However, in general, these four terms are not ignorable. Therefore, Li and Chen's procedure is not suitable for temporal dependent data even in the two-sample case.

We now study the asymptotic distribution of $\hat{D}_{nt}$. The following theorem establishes the asymptotic normality of $\hat{D}_{nt}$. The proof is given in Section 2.7.

Theorem 2. *Under Conditions 1–2,* $\sigma_{nt}^{-1}(\hat{D}_{nt} - D_t) \to N(0,1)$ *in distribution as* $n \to \infty$, *where* σ_{nt}^2 *is defined in Theorem 1.*

We do not require explicit conditions on p and n in Theorem 2. The asymptotic normality holds provided Conditions 1–2 hold. In particular, we only need Condition 1 under the null hypothesis. Thus, our test is valid under Condition 1 without Condition 2, which is needed only for studying the power of the test. The normality assumption in model (2.3) is not essential and can be relaxed to a multivariate model as considered in Chen et al. (2010) and Li and Chen (2012). See Subsection 2.3.1 for the generalization to the non-Gaussian case.

Under H_0, $D_t = 0$ for all $t \in \{1, \ldots, T-1\}$. Theorem 2 indicates that $\sigma_{nt,0}^{-1}\hat{D}_{nt}$ converges to $N(0,1)$ in distribution where $\sigma_{nt,0}^2 = 4V_{0t}/\{nw(t)\}^2$ is the variance of $\hat{D}_{nt}$ under H_0. An asymptotic α-level rejection region is $R_t = \{\sigma_{nt,0}^{-1}\hat{D}_{nt} > z_\alpha\}$, where z_α is the upper α quantile of the standard normal distribution. For each $t \in \{1, \ldots, T-1\}$, one can use R_t to test for the hypothesis in (2.1). Provided that one test based on $\hat{D}_{nt}$ rejects the null hypothesis, one may suspect that change points could exist among covariance matrices. Accordingly, t, in $\hat{D}_{nt}$, could be considered as a tuning parameter, and it is hard to decide which t should be used for testing in practice. To make the proposed method free of any tuning parameter and adaptive to unknown change points, we propose the following statistic for testing the hypothesis in (2.1):

$$\mathcal{M}_n = \max_{1 \le t \le T-1} \hat{\sigma}_{nt,0}^{-1}\hat{D}_{nt}, \tag{2.6}$$

where $\hat{\sigma}_{nt,0}^2 = 4\hat{V}_{0t}/\{nw(t)\}^2$. The estimator $\hat{V}_{0t}$ can be constructed by replacing

$\mathrm{tr}(C_{s_uh_k}C^{\mathrm{T}}_{s_vh_l})$ in V_{0t} with $U_{s_us_v,h_kh_l}$, an unbiased estimator of $\mathrm{tr}(C_{s_uh_k}C^{\mathrm{T}}_{s_vh_l})$. Define $U_{s_us_v,h_kh_l} = U_{s_us_v,h_kh_l,0} - U_{s_us_v,h_kh_l,1} - U_{s_vs_u,h_lh_k,1} + U_{s_us_v,h_kh_l,2}$, where $U_{s_us_v,h_kh_l,0} = (1/P_n^2)\sum_{i\neq j=1}^{n} Y^{\mathrm{T}}_{is_u}Y_{js_v}Y^{\mathrm{T}}_{ih_k}Y_{jh_l}$ is an unbiased estimator of $\mathrm{tr}(C_{s_uh_k}C^{\mathrm{T}}_{s_vh_l}) + \mu^{\mathrm{T}}_{s_v}C_{s_uh_k}\mu_{h_l} + \mu^{\mathrm{T}}_{s_u}C_{s_vh_l}\mu_{h_k} + \mu^{\mathrm{T}}_{s_u}\mu_{s_v}\mu^{\mathrm{T}}_{h_k}\mu_{h_l}$, $U_{s_us_v,h_kh_l,1} = (1/P_n^3)\tilde{\sum}_{i,j,g} Y^{\mathrm{T}}_{is_u}Y_{js_v}Y^{\mathrm{T}}_{ih_k}Y_{gh_l}$ is an unbiased estimator of $\mu^{\mathrm{T}}_{s_v}C_{s_uh_k}\mu_{h_l} + \mu^{\mathrm{T}}_{s_u}\mu_{s_v}\mu^{\mathrm{T}}_{h_k}\mu_{h_l}$ and an unbiased estimator of $\mu^{\mathrm{T}}_{s_u}\mu_{s_v}\mu^{\mathrm{T}}_{h_k}\mu_{h_l}$ is $U_{s_us_v,h_kh_l,2} = (1/P_n^4)\tilde{\sum}_{i,j,g,f} Y^{\mathrm{T}}_{is_u}Y_{js_v}Y^{\mathrm{T}}_{gh_k}Y_{fh_l}$. A computation efficient formulation of the estimators $U_{s_us_v,h_kh_l,q}$, $q = 1, 2$, is similar to that for $U_{s_1,s_2,q}$ defined in (2.4).

Under H_0 and Condition 1, similar to the derivation in Lemma 4 in Section 2.7, the leading order of the $\mathrm{cov}(\hat{D}_{nt}, \hat{D}_{nq})$ is $Q_{n,tq}$, where

$$Q_{n,tq} = \sum_{s_1=1}^{t}\sum_{h_1=1}^{q}\sum_{s_2=t+1}^{T}\sum_{h_2=q+1}^{T} V_{n0}(s_1, s_2, h_1, h_2)/\{w(t)w(q)\}$$

and $V_{n0}(s_1, s_2, h_1, h_2) = (4/n^2)\sum_{u,v,k,l\in\{1,2\}}(-1)^{|u-v|+|k-l|}\mathrm{tr}^2(C_{s_uh_k}C^{\mathrm{T}}_{s_vh_l})$. Then the covariance between $\sigma^{-1}_{nq,0}\hat{D}_{nt}$ and $\sigma^{-1}_{nq,0}\hat{D}_{nq}$ is $Q_{n,ts}/\sqrt{(Q_{n,tt}Q_{n,ss})}$, which is the correlation between $\hat{D}_{nt}$ and $\hat{D}_{nq}$.

Let V_{nD} be a correlation matrix whose (t, s) component is $Q_{n,ts}/\sqrt{(Q_{n,tt}Q_{n,ss})}$ for $t, s \in \{1, \ldots, T-1\}$. Assume that V_{nD} converges to V_D as $n \to \infty$. The following theorem provides the asymptotic distribution of $\mathcal{M}_n$.

Theorem 3. *Under Condition 1, we have that under H_0, $\mathcal{M}_n \to W$ in distribution as $n \to \infty$, where $W = \max_{1\leq t\leq T-1} Z_t$ and $Z = (Z_1, \ldots, Z_{T-1})^{\mathrm{T}}$ is a multivariate normally distributed random vector with mean 0 and covariance V_D.*

According to Theorem 3, an α-level test for (2.1) rejects the null hypothesis if $\mathcal{M}_n > W_\alpha$, where W_α is the α-quantile of W such that $\mathrm{pr}(W > W_\alpha) = \alpha$. Let Z_n be a $N(0, \hat{V}_{nD})$ distributed random vector with the (t, s) component of $\hat{V}_{nD}$ estimated by $\hat{Q}_{n,ts}/\sqrt{(\hat{Q}_{n,tt}\hat{Q}_{n,ss})}$, where

$$\hat{Q}_{n,ts} = \frac{4}{n^2 w(t)w(s)}\sum_{s_1=1}^{t}\sum_{h_1=1}^{s}\sum_{s_2=t+1}^{T}\sum_{h_2=s+1}^{T}\sum_{u,v,k,l\in\{1,2\}}(-1)^{|u-v|+|k-l|}U^2_{s_us_v,h_kh_l},$$

and $U_{s_u s_v, h_k h_l}$ is defined just below (2.6). Simulations suggest that the plug-in estimates of the correlation matrix $\hat{V}_{nD}$ are reliable when the sample size is approximate 40 or above. See Section 2.7 for a detailed comparison between $\hat{V}_{nD}$ and V_{nD}. The quantile W_α can be approximated by $W_{n,\alpha}$ obtained from the multivariate normal distribution by finding the quantile $w_{n,\alpha} = (W_{n,\alpha}, \ldots, W_{n,\alpha})^{\mathrm{T}}$ satisfying $\mathrm{pr}(Z_n < w_{n,\alpha}) = 1 - \alpha$. The quantile $w_{n,\alpha}$ can be computed using the R package mvtnorm (Genz et al., 2018), and no simulation is needed to find quantile $W_{n,\alpha}$.

The lower bound for power based on $\mathcal{M}_n$ is

$$\mathrm{pr}(\mathcal{M}_n > W_\alpha) \geq \max_{1 \leq t \leq T-1} \mathrm{pr}(\hat{\sigma}_{nt,0}^{-1} \hat{D}_{nt} > W_\alpha) = \max_{1 \leq t \leq T-1} \Phi\Big(- \frac{\sigma_{nt,0}}{\sigma_{nt}} W_\alpha + \frac{D_t}{\sigma_{nt}} \Big), \quad (2.7)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. If D_t/σ_{nt} dominates W_α, the right-hand side of (2.7) is the maximum power of the test using R_t constructed on a single $\hat{D}_{nt}$, so the test based on $\mathcal{M}_n$ is more powerful than any test based on a single $\hat{D}_{nt}$.

2.3.1 Non-Gaussian random errors

To relax the Gaussian assumption, we assume the following data generation model for $\varepsilon_i = (\varepsilon_{i1}^{\mathrm{T}}, \ldots, \varepsilon_{iT}^{\mathrm{T}})^{\mathrm{T}}$ and $\varepsilon_i = \Gamma Z_i$ where $\Gamma = (\Gamma_1^{\mathrm{T}}, \ldots, \Gamma_T^{\mathrm{T}})^{\mathrm{T}}$ is a $Tp \times m$ matrix with $m \geq Tp$ such that $\Sigma = \Gamma \Gamma^{\mathrm{T}}$ and $\Gamma_s \Gamma_t^{\mathrm{T}} = C_{st}$. We assume $Z_1, \ldots, Z_n$ are independent and identically distributed m-dimensional random vectors such that $E(Z_1) = 0$ and $\mathrm{var}(Z_1) = I_m$. Write $Z_1 = (Z_{11}, ..., Z_{1m})^{\mathrm{T}}$. We assume that each Z_{1l} has a uniformly bounded 8th moment. Also, we assume there exists a finite constant such that for $l = 1, \ldots, m$, $E(Z_{1l}^4) = 3 + \Delta$ and for any integers $l_v \geq 0$ with $\sum_{v=1}^{q} l_v = 8$, $E(Z_{1i_1}^{l_1} \cdots Z_{1i_q}^{l_q}) = E(Z_{1i_1}^{l_1}) \cdots E(Z_{1i_q}^{l_q})$, whenever $i_1, \ldots, i_q$ are distinct indices.

Under Condition 1 and the above setup, it can be shown that the leading order of the variance of $\hat{D}_{nt}$ is

$$\begin{aligned}\operatorname{var}(\hat{D}_{nt}) = {} & \frac{4}{n^2 w^2(t)} \sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}} \sum_{\substack{u,v,\\ k,l\in\{1,2\}}} (-1)^{|u-v|+|k-l|} \operatorname{tr}^2(C_{s_u h_k} C^{\mathrm{T}}_{s_v h_l}) \\ & + \frac{8}{n w^2(t)} \sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}} \sum_{u,k\in\{1,2\}} (-1)^{|u-k|} \big[\operatorname{tr}\{(\Sigma_{s_1} - \Sigma_{s_2}) C_{s_u h_k} (\Sigma_{h_1} - \Sigma_{h_2}) C^{\mathrm{T}}_{s_u h_k}\} \\ & \qquad + \Delta \operatorname{tr}\{\Gamma^{\mathrm{T}}_{s_u} (\Sigma_{s_1} - \Sigma_{s_2}) \Gamma_{s_u} \circ \Gamma^{\mathrm{T}}_{h_k} (\Sigma_{h_1} - \Sigma_{h_2}) \Gamma_{h_k}\}\big].\end{aligned}$$

Under the null hypothesis, $\operatorname{var}(\hat{D}_{nt}) = 4V_{0t}/\{n^2 w^2(t)\}$. The variance V_{0t} can be estimated using the formula given below equation (2.6). The results in Theorems 2 and 3 can be established in a similar way.

2.3.2 Power-enhanced test for sparse alternatives

The proposed test statistic, $\mathcal{M}_n$, is powerful for alternatives with small absolute differences in many components of Σ_t. However, it might not be very powerful for sparse alternatives with the differences among Σ_t only residing in a few components. To enhance the power of the proposed test for sparse alternatives, we include an additional term with $\mathcal{M}_n$, as an idea in Fan et al. (2015).

Let $\bar{Y}_{s_1 v} = \sum_{i=1}^{n} Y_{i s_1 v}/n$ be the sample mean of the vth component measured at time s_1, and define $\hat{\sigma}_{s_1,uv} = \sum_{i=1}^{n} (Y_{i s_1 u} - \bar{Y}_{s_1 u})(Y_{i s_1 v} - \bar{Y}_{s_1 v})/(n-1)$ as the sample covariance between components $u, v \in \{1, \ldots, p\}$ at time s_1. Define $\hat{D}_{nt,uv} = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} (\hat{\sigma}_{s_1,uv} - \hat{\sigma}_{s_2,uv})^2$ as an estimator of $D_{nt,uv} = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} (\sigma_{s_1,uv} - \sigma_{s_2,uv})^2$. The estimator $\hat{D}_{nt,uv}$ is a consistent estimator of $D_{nt,uv}$. Let $C^{(uv)}_{s_k h_t}$ be the (u, v) component of $C_{s_k h_t}$ and $\sigma^{(uv)}_{h_t}$ is the (u, v) component of Σ_{h_t}. To define the variance of $\hat{D}_{nt,uv}$, define the following notation:

$$\begin{aligned} F^{(uv)}_{s_k s_l h_t h_s} &= \sigma^{(uv)}_{s_l}\sigma^{(uv)}_{h_t}\{C^{(vu)}_{h_s s_k}C^{(uv)}_{s_k h_s} + C^{(vv)}_{h_s s_k}C^{(uu)}_{s_k h_s}\} \\ &+ \sigma^{(uv)}_{s_l}\sigma^{(uv)}_{h_s}\{C^{(vu)}_{h_t s_k}C^{(uv)}_{s_k h_t} + C^{(vv)}_{h_t s_k}C^{(uu)}_{s_k h_t}\} \\ &+ \sigma^{(uv)}_{s_k}\sigma^{(uv)}_{h_t}\{C^{(vu)}_{h_s s_l}C^{(uv)}_{s_l h_s} + C^{(vv)}_{h_s s_l}C^{(uu)}_{s_l h_s}\} \\ &+ \sigma^{(uv)}_{s_k}\sigma^{(uv)}_{h_s}\{C^{(vu)}_{h_t s_l}C^{(uv)}_{s_l h_t} + C^{(vv)}_{h_t s_l}C^{(uu)}_{s_l h_t}\}, \end{aligned}$$

$$\begin{aligned} G^{(uv)}_{s_k s_l h_t h_s} &= \{C^{(vu)}_{s_k h_s}C^{(uv)}_{s_k h_s} + C^{(vv)}_{s_k h_s}C^{(uu)}_{s_k h_s}\}\{C^{(vu)}_{s_l h_t}C^{(uv)}_{s_l h_t} + C^{(vv)}_{s_l h_t}C^{(uu)}_{s_l h_t}\} \\ &+ \{C^{(vu)}_{s_k h_t}C^{(uv)}_{s_k h_t} + C^{(vv)}_{s_k h_t}C^{(uu)}_{s_k h_t}\}\{C^{(vu)}_{s_l h_s}C^{(uv)}_{s_l h_s} + C^{(vv)}_{s_l h_s}C^{(uu)}_{s_l h_s}\}. \end{aligned}$$

The leading order term of the variance of $\hat{D}_{nt,uv}$ is

$$\sigma^2_{nt,uv} = \frac{1}{w^2(t)} \sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}} \sum_{\substack{k,l,\\ s,t\in\{1,2\}}} (-1)^{|k-l|+|s-t|}\{n^{-1}F^{(uv)}_{s_k s_l h_t h_s} + n^{-2}G^{(uv)}_{s_k s_l h_t h_s}\}. \tag{2.8}$$

Under H_0, the first term in (2.8) is 0. Namely,

$$\sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}} \sum_{\substack{k,l,\\ s,t\in\{1,2\}}} (-1)^{|k-l|+|s-t|}F^{(uv)}_{s_k s_l h_t h_s} = 0.$$

The leading term in the variance of $\hat{D}_{nt,uv}$ under H_0 is

$$\sigma^2_{nt,uv0} = \sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}} \sum_{\substack{k,l,\\ s,t\in\{1,2\}}} (-1)^{|k-l|+|s-t|}G^{(uv)}_{s_k s_l h_t h_s}/n^2.$$

Let $\hat{G}^{(uv)}_{s_k s_l h_t h_s}$ be a sample plug-in estimate of $G^{(uv)}_{s_k s_l h_t h_s}$, and $\hat{\sigma}^2_{nt,uv0}$ be the corresponding sample estimate of $\sigma^2_{nt,uv0}$. Then, the power-enhanced test statistic is

$$\mathcal{M}^*_n = \max_{1\le t\le T-1}\Big\{\hat{\sigma}^{-1}_{nt,0}\hat{D}_{nt} + \lambda_n \sum_{u\le v} I(\hat{D}_{nt,uv} > \delta_{n,p}\hat{\sigma}_{nt,uv0})\Big\},$$

where $\delta_{n,p}$ and λ_n are tuning parameters. The tuning parameters are chosen such that the second part of $\mathcal{M}^*_n$ equals zero with probability tending to one under H_0, and it converges to a large number under sparse alternatives.

We now discuss the choices for tuning parameters for the above power-enhanced test statistic. Let $R = (\rho_{ij})$ be the correlation matrix corresponding to the common covariance Σ_1 under H_0. Define $N_j(\alpha) = \text{card}\{i : |\rho_{ij}| > (\log p)^{-1-\alpha}\}$ and $\Lambda(r) = \{i : |\rho_{ij}| > r \text{ for some } j \neq i\}$. We assume the following condition used in Cai et al. (2013).

Condition 3. *Suppose that there exists a* α *and a set* $\pi \subset \{1, \ldots, p\}$ *whose size is* $o(p)$ *such that* $\max_{1\le j\le p, j\notin\pi} N_j(\alpha) = o(p^\gamma)$ *for all* $\gamma > 0$. *In addition, there exists a* $r < 1$ *and a sequence of numbers* $\Lambda_{p,r} = o(p)$ *so that card*$\{\Lambda(r)\} \le \Lambda_{p,r}$.

Define $l_{s_1 s_2} = \max_{1\le u\le v\le p}(\hat\sigma_{s_1,uv} - \hat\sigma_{s_2,uv})^2/\sigma_{ns_1s_2,uv0}$ where $\sigma^2_{ns_1s_2,uv0} = \text{var}\{(\hat\sigma_{s_1,uv} - \hat\sigma_{s_2,uv})^2\}$ under H_0. Similar to the proof of Theorem 1 in Cai et al. (2013), under Condition 3 and H_0, we can show that

$$\text{pr}\{l_{s_1s_2} - 4\log(p) + \log\log(p) \le t\} \to \exp\{-\exp(-t/2)/\sqrt{(8\pi)}\}. \tag{2.9}$$

Define $L_{uv} = \hat D_{nt,uv}/\hat\sigma_{nt,uv0}$ and $L_n = \max_{1\le u\le v\le p} L_{uv}$. Denote the second term in M_n^* as $M_{n1}^* = \lambda_n \sum_{u\le v} I(\hat D_{nt,uv} > \delta_{n,p}\hat\sigma_{nt,uv0})$. Because $\sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\sigma_{ns_1s_2,uv0}/\sigma_{nt,uv0} \le K$, uniformly for all u, v for a constant $K > 0$, and uniform consistency of $\hat\sigma_{nt,uv0}$ to $\sigma_{nt,uv0}$, we have, under H_0,

$$\begin{aligned}
\text{pr}(M_{n1}^* = 0) &\ge \text{pr}\Big(L_n \le \delta_{n,p}) = \text{pr}(\max_{1\le u\le v\le p} \hat D_{nt,uv}/\hat\sigma_{nt,uv0} \le \delta_{n,p}\Big) \\
&= \text{pr}\Big\{\max_{1\le u\le v\le p} \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T} \frac{(\hat\sigma_{s_1,uv} - \hat\sigma_{s_2,uv})^2}{\sigma_{ns_1s_2,uv0}} \frac{\sigma_{ns_1s_2,uv0}}{\sigma_{nt,uv0}} \le \delta_{n,p}\Big\} \\
&\ge \text{pr}\Big\{\max_{1\le u\le v\le p} \max_{\substack{1\le s_1\le t,\\ t+1\le s_2\le T}} \frac{(\hat\sigma_{s_1,uv} - \hat\sigma_{s_2,uv})^2}{\sigma_{ns_1s_2,uv0}} \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T} \frac{\sigma_{ns_1s_2,uv0}}{\sigma_{nt,uv0}} \le \delta_{n,p}\Big\} \\
&\ge \text{pr}\Big\{\max_{1\le u\le v\le p} \max_{\substack{1\le s_1\le t,\\ t+1\le s_2\le T}} (\hat\sigma_{s_1,uv} - \hat\sigma_{s_2,uv})^2/\sigma_{ns_1s_2,uv0} \le \delta_{n,p}/K\Big\} \\
&\ge 1 - \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T} \text{pr}\Big(l_{s_1s_2} > \delta_{n,p}/K\Big).
\end{aligned}$$

Applying the result in (2.9), if $\delta_{n,p}/K - 4\log(p) + \log\log(p) \to \infty$, then $\text{pr}(M_{n1}^* = 0) \to 1$. We suggest choose $\delta_{n,p}$ at the order of $\log(n)\log(p)$ and λ_n to be a constant based on our numerical experiments. In summary, the tuning parameters $\delta_{n,p}$ and λ_n ensure that, under the null hypothesis, M^*_{n1} converges to zero with probability one.

2.4 Change point identification

If H_0 is rejected, then there exist change points among the covariances Σ_t. We first consider an alternative with one change point:

$$H_1^*: \quad \Sigma_1 = \cdots = \Sigma_{k_1} \neq \Sigma_{k_1+1} = \cdots = \Sigma_T, \tag{2.10}$$

where k_1 is the true change point, whose location is estimated by

$$\hat{k}_1 = \arg \max_{1 \le t \le T-1} \hat{D}_{nt}. \tag{2.11}$$

Define the weight function

$$r(t;k) = \begin{cases} (T-k)/(T-t), & 1 \le t \le k, \\ k/t, & k+1 \le t \le T-1. \end{cases}$$

For any fixed value $k \in \{1, \ldots, T-1\}$, the function $r(t;k)$ achieves its maximum value at $t = k$. Let $\beta_n = \max_{1 \le t \le T-1} \max\{\sqrt{V_{0t}}, \sqrt{(nV_{1t})}\}$ and $\Delta_n = \operatorname{tr}\{(\Sigma_1 - \Sigma_T)^2\}$. The next theorem establishes the rate of convergence of the change point estimator $\hat{k}_1$ obtained by (2.11) under the alternative H_1^*.

Theorem 4. *Under the alternative H_1^* in (2.10), $E(\hat{D}_{nt}) = D_t = r(t;k_1)\Delta_n$ and D_t attains its maximum at $t = k_1$. Moreover, $\hat{k}_1 - k_1 = O_p\{\beta_n/(n\Delta_n)\}$.*

Since $r(t;k_1)$ achieves its maximum at $t = k_1$, the first part of Theorem 4 indicates that $t = k_1$ maximizes $E(\hat{D}_{nt})$ as a function of t. This is the rationale for estimating k_1 through (2.11). When the data dimension is fixed, $\hat{k}_1 - k_1 = O_p(1/\sqrt{n})$. The effect of data dimension is reflected both in β_n and Δ_n. Here β_n can be considered as noise and Δ_n can be viewed as signal. If the signal level is larger than the noise level, the rate of convergence of $\hat{k}_1$ is faster than $O_p(1/\sqrt{n})$. On the other hand, if β_n is not smaller than $n\Delta_n$, $\hat{k}_1$ is not consistent.

Next, we consider the alternative, H_1, with multiple change points $k_1 < \cdots < k_q$, as specified in (2.2). Under H_1^*, we have shown in Theorem 4 that the maximum of D_t is attained at change point k_1.

Theorem 5. *Under H_1 in (2.2), the maximum value of D_t is attained at one of the change points among $k_1 < \cdots < k_q$.*

If we estimate the multiple change points by repeatedly applying estimation methods in (2.11) to the population version D_t to all sub-sequences with non-zero D_t, Theorem 5 ensures that we find all the true change points. This property is important for applying the binary segmentation method to identify multiple change points as demonstrated by Venkatraman (1992) in an unpublished technical report.

To describe the proposed binary segmentation method, we first define some notation. Let $[I_t]$ represent the quantities computed based on the data within the time interval I_t, a subset of $[1, T]$. For example, $\mathcal{M}_n[t_1, t_2]$ is the test statistic defined in (2.6) calculated based on $Y[t_1, t_2]$, the data collected between time $t = t_1$ and $t = t_2$ for $t_1 < t_2$. Namely, $\mathcal{M}_n[t_1, t_2] = \max_{t_1 \le t < t_2} \hat{\sigma}_{nt,0}^{-1}[t_1, t_2]\hat{D}_{nt}[t_1, t_2]$.

The binary segmentation method can be summarized as follows. Let α_n be a number specified in Theorem 6. In the first step, compute $\mathcal{M}_n[1, T]$. If $\mathcal{M}_n[1, T] < W_{\alpha_n}[1, T]$, where $W_{\alpha_n}[1, T]$ is the cut-off quantile estimated based on $Y[1, T]$, we accept the null hypothesis and stop. Otherwise, we identify the change point, say $\hat{k}_1$, using (2.11). Next, we compute $\mathcal{M}_n$ for both subsequences $Y[1, \hat{k}_1]$ and $Y[\hat{k}_1 + 1, T]$. For each subsequence, we repeat the first step until no change points can be identified or the number of repeated measurements in the subsequence is less than two.

Let I_t be any interval of the form $[k_f + 1, k_g]$ with $f + 1 < g$ where $f \in \{0, \ldots, q - 1\}$ and $g \in \{2, \ldots, q + 1\}$ that contains at least one change point k_j for $j \in \{1, \ldots, q\}$, where $k_0 = 0$ and $k_{q+1} = T$. Denote $\text{mSNR} = \min_{I_t} \max_{k_s \in I_t} \sigma_{nk_s,0}^{-1}[I_t]D_{k_s}[I_t]$ as the smallest maximum signal-to-noise ratio among all segmentations I_t.

Theorem 6. *Assume* $\alpha_n \to 0$ *and* $\mathrm{m}SNR$ *diverges so that* $W_{\alpha_n} = o(\mathrm{m}SNR)$. *For all* I_t, *if* $\beta_n[I_t] = o(nD_{k_s}[I_t])$ *for some change points* $k_s \in I_t$, *we have,* $\lim_{n\to\infty} \mathrm{pr}(\hat{q} = q; \hat{k}_j = k_j,\ j = 1, \ldots, q) = 1$.

The first assumption, with $W_{\alpha_n} = o(\mathrm{m}SNR)$, is a very mild condition, which ensures the consistency of the proposed test at each step of the binary segmentation. The second assumption $\beta_n[I_t] = o(nD_{k_s}[I_t])$ is needed to ensure the consistency of the change point estimators. Theorem 6 implies that the proposed binary segmentation procedure consistently estimates the number and locations of the change points.

2.5 Simulation studies

In this section, we present multiple simulation studies to demonstrate the finite sample performance of the proposed method. The data were generated from the model

$$Y_{it} = \mu_t + \sum_{h=0}^{L} A_{t,h}\eta_{i(t-h)} \quad (i = 1, \ldots, n;\ t = 1, \ldots, T),$$

where $A_{t,h}$ is a $p \times p$ matrix, $\mu_t = 0$ and η_{it} are p-dimensional multivariate normally distributed random vectors with mean 0 and covariance I_p. Let $t \geq s$. This implies that $\mathrm{cov}(Y_{it}, Y_{is}) = \sum_{h=t-s}^{L} A_{t,h}A_{s,h-(t-s)}^{\mathrm{T}}$ if $t - s \leq L$ and $\mathrm{cov}(Y_{it}, Y_{is}) = 0$ if $t - s > L$, and allows dependence among components within the vector Y_{it} and dependence among $\{Y_{it}\}_{t=1}^{T}$ at different time points. In the simulation studies, we set $n = 40, 50$ and 60, $p = 500, 750$ and 1000, $T, = 5$ and 8, and $L = 3$. The simulation results reported in Tables 2.1 and 2.2 were based on 500 replications. The results in Table 2.3 were based on 100 simulation replications.

Let $k_1 = [T/2]$ be the largest integer no greater than $T/2$. For $t \in \{1, \ldots, k_1\}$, we set $A_{t,h} = A^{(1)}$ for $h \in \{0, \ldots, L\}$. For $t \in \{k_1 + 1, \ldots, T\}$ and $h \in \{0, \ldots, L\}$, $A_{t,h} = A^{(2)}$. Two simulation settings were used for the generation of the A matrices. In setting (I), we set $A^{(1)} = \big\{0.6^{|i-j|} I(|i - j| < p/5)\big\}$, and $A^{(2)} = \big\{(0.6 + \delta)^{|i-j|} I(|i - j| < p/5)\big\}$. If $\delta = 0$, $A^{(1)}$ and $A^{(2)}$ are the same and the covariances of Y_{it} are the same for all t.

Hence, the null hypothesis, H_0, is true. If $\delta \neq 0$, the null hypothesis is false and k_1 is the true change point. In setting (II), we set $A^{(1)} = \{(|i-j|+1)^{-2} I(|i-j| < p/5)\}$ and $A^{(2)} = \{(|i-j|+\delta+1)^{-2} I(|i-j| < p/5)\}$. Similar to setting (I), a value of $\delta = 0$ corresponds to the null hypothesis being true. If $\delta \neq 0$, k_1 is the underlying true change point for the covariance matrices.

Table 2.1 demonstrates the empirical size and power of the proposed test for the homogeneity of covariance matrices under setting (I) at nominal level 0.05. We observe that the size of the proposed test is reasonably close to the nominal level. The power increases as n increases, as δ increases, and as T increases. Table 2.1 also provides the empirical size and power of the proposed test under simulation setting (II). The phenomena in setting (II) are very similar to those in setting (I).

Table 2.1: Empirical size and power of the proposed test, percentages of simulation replications that reject the null hypothesis under settings (I) and (II)

			$T=5$			$T=8$		
			p			p		
Setting	δ	n	500	750	1000	500	750	1000
(I)	0(size)	40	4.6	4.8	6.4	4.8	4.8	4.4
		50	4.6	5.2	5.4	4.4	5.8	4.6
		60	6.0	4.4	4.2	5.4	4.2	3.6
	0.05	40	21.4	27.6	24.8	35.6	34.6	34.2
		50	37.0	36.0	36.0	49.8	48.8	52.0
		60	45.6	49.2	46.2	59.6	65.6	65.0
	0.10	40	99.6	100	99.8	100	100	100
		50	100	100	100	100	100	100
		60	100	100	100	100	100	100
(II)	0(size)	40	4.4	5.4	5.0	4.4	4.0	4.8
		50	5.6	4.6	4.8	6.0	5.2	5.6
		60	4.8	4.6	4.2	3.6	5.6	5.0
	0.10	40	33.4	35.8	38.2	50.2	52.0	51.6
		50	44.2	48.6	47.0	68.4	70.6	74.0
		60	65.4	63.6	60.4	87.0	89.6	88.0
	0.20	40	99.8	99.8	99.6	100	100	100
		50	99.8	100	100	100	100	100
		60	100	100	100	100	100	100

The percentages of correct identification are summarized in Table 2.2 when the null hypothesis is false under settings (I) and (II). The percentages of correct identification are the percentages of simulation replications that estimate the location of the change point correctly among all those that reject the null hypothesis. When $T = 5$, the true change point is $k_1 = 2$, and when $T = 8$, the true change point is $k_1 = 4$. In both settings, for almost all the cases, the percentages increase as n and δ increase.

Table 2.2: Percentages of correct change point identification among all rejected hypotheses under settings (I) and (II)

				$T = 5$			$T = 8$	
				p			p	
Setting	δ	n	500	750	1000	500	750	1000
(I)	0.05	40	41.12	37.96	40.65	30.18	29.88	37.58
		50	51.35	45.81	43.33	39.52	39.34	41.54
		60	52.63	53.28	52.17	49.33	49.70	55.08
	0.10	40	93.17	96.60	95.19	93.79	93.80	96.40
		50	98.00	98.60	98.20	98.40	97.20	99.00
		60	99.20	99.80	99.40	99.80	98.60	99.00
(II)	0.10	40	49.10	45.51	55.50	43.12	47.15	47.10
		50	65.00	61.51	55.98	53.80	61.19	58.65
		60	72.78	69.72	64.57	67.59	72.99	75.23
	0.20	40	90.58	90.98	89.16	95.80	96.00	95.60
		50	93.37	92.60	93.20	98.60	98.20	99.40
		60	97.00	96.20	96.40	99.80	99.80	99.80

To demonstrate the performance of the proposed binary segmentation procedure for identifying multiple change points, we generated data using simulation setup (II) with two change points, k_1 and k_2. When $T = 5$, $k_1 = 2$ and $k_2 = 4$. When $T = 8$, $k_1 = 4$ and $k_2 = 6$. For $t \in \{k_{j-1}+1, \ldots, k_j\}$, we set $A_{t,h} = A^{(j)}$ for $h \in \{0, \ldots, L\}$ and $j = 1, 2, 3$ with $k_0 = 0$ and $k_3 = T$. Here, $A^{(1)}$ and $A^{(2)}$ were set to be the same as those in setting (II), and we set $A^{(3)} = A^{(1)}$. The values of δ were chosen to be 0.15 and 0.25. The average true positives and the average true negatives are summarized in Table 2.3. The true positives are the correctly-identified change points, and the true negatives are the correctly-identified time points where no covariance change exists. For $T = 5$, the maximum number of true positives and true negatives for each is 2. For $T = 8$, the maximum number of true positives and true negatives is 2 and 5, respectively. The results in Table 2.3 show that the proposed binary segmentation procedure performs well as the sample size, n, increases and as the signal, δ, increases.

Table 2.3: Average true positives and average true negatives for identifying multiple change points using the proposed binary segmentation method. Standard errors are included after each number. For $T = 5$, the maximum number of true positives and true negatives for each is 2. For $T = 8$, the maximum number of true positives and true negatives is 2 and 5, respectively

			δ=0.15				δ=0.25			
T	p	n	ATP	SE	ATN	SE	ATP	SE	ATN	SE
5	500	40	1.10	0.36	1.90	0.30	1.81	0.39	1.92	0.27
		50	1.36	0.48	1.87	0.37	1.94	0.24	1.98	0.14
		60	1.57	0.50	1.92	0.27	2.00	0.00	1.92	0.28
	750	40	1.11	0.37	1.82	0.41	1.76	0.43	1.94	0.24
		50	1.38	0.49	1.92	0.27	2.00	0.00	1.96	0.24
		60	1.47	0.50	1.90	0.30	2.00	0.00	1.98	0.14
	1000	40	1.15	0.36	1.90	0.30	1.87	0.34	1.95	0.22
		50	1.22	0.42	1.96	0.20	1.96	0.20	1.96	0.20
		60	1.54	0.50	1.96	0.20	2.00	0.00	1.98	0.14
8	500	40	1.40	0.49	4.84	0.40	1.91	0.29	4.90	0.30
		50	1.62	0.49	4.85	0.36	1.97	0.17	4.92	0.27
		60	1.78	0.42	4.89	0.32	2.00	0.00	4.95	0.22
	750	40	1.52	0.50	4.82	0.39	1.90	0.30	4.85	0.36
		50	1.67	0.47	4.83	0.38	1.97	0.17	4.94	0.24
		60	1.81	0.40	4.90	0.34	2.00	0.00	4.90	0.30
	1000	40	1.43	0.50	4.82	0.44	1.88	0.33	4.92	0.27
		50	1.68	0.47	4.80	0.40	1.99	0.10	4.96	0.25
		60	1.84	0.37	4.92	0.27	2.00	0.00	4.94	0.24

2.5.1 Power-enhanced test statistic

We conducted a numerical simulation to illustrate the performance of the power-enhanced test statistic under sparse alternatives. The data were generated according to setting (I), except for a sparse alternative design. Specifically, let $k_1 = [T/2]$ be the largest integer no greater than $T/2$. For $t \in \{1, \ldots, k_1\}$, we set $A_{t,h} = A^{(1)}$ for $h \in \{0, \ldots, L\}$. For $t \in \{k_1 + 1, \ldots, T\}$, we set $A_{t,h} = A^{(2)}$, where $A_h^{(1)} = \left\{0.6^{|i-j|} I(|i-j| < p/5)\right\}$. Under the null hypothesis, $A_h^{(2)}$ was set equal to $A_h^{(1)}$. Under the sparse alternative hypothesis, $A_h^{(2)}$ was the same as $A_h^{(1)}$ except the components within $\{|i-j| < 2, i < p/25\}$ were set to 1.4.

Table 2.4: Empirical size and power, percentages of simulation replications that reject the null hypothesis for the test statistic $\mathcal{M}_n$ and the power-enhanced test statistic $\mathcal{M}_n^*$

		$\mathcal{M}_n$		$\mathcal{M}_n^*$	
n	p	Null	Alternative	Null	Alternative
40	500	5.2	35.4	6.0	67.6
40	750	3.2	34.6	3.6	62.8
40	1,000	4.6	36.4	4.6	62.4
50	500	5.2	47.8	5.6	91.6
50	750	6.4	47.2	6.6	94.2
50	1,000	3.4	52.6	3.4	97.0
60	500	3.8	56.8	4.4	98.8
60	750	4.8	65.8	5.6	99.2
60	1,000	4.2	66.4	4.2	99.8
80	500	4.0	81.2	4.0	100
80	750	3.2	86.4	3.8	100
80	1,000	4.6	84.8	4.6	100

Table 2.4 reports the empirical size and power of the test based on $\mathcal{M}_n$ and $\mathcal{M}_n^*$. In the simulation, the tuning parameter $\delta_{n,p}$ was set to $0.5\log(n)\log(p)$, and λ_n was set to 0.15. We observe that both tests can control the type I error, and the power-enhanced test does not inflate the type I error. More importantly, the power-enhanced test statistic has greater power under the sparse alternative setting.

2.5.2 Non-Gaussian random errors

To illustrate the numerical performance of the proposed method under the non-Gaussian setup, we generated data from the linear process model $Y_{it} = \mu_t + \sum_{h=0}^{L} A_{t,h}\eta_{i(t-h)}$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $A_{t,h}$ is a $p \times p$ matrix, $\mu_t = 0$ and η_{it} are p-dimensional random vectors with each element independently generated from a standardized Gamma distribution with shape parameter 4 and scale parameter 0.5.

Let $k_1 = [T/2]$ be the largest integer no greater than $T/2$. For $t \in \{1, \ldots, k_1\}$, we set $A_{t,h} = A^{(1)} = \{0.6^{|i-j|} I(|i-j| < p/5)\}$. For $t \in \{k_1 + 1, \ldots, T\}$, we set $A_{t,h} = A^{(2)} = \{(0.6+\delta)^{|i-j|} I(|i-j| < p/5)\}$. If $\delta = 0$, $A^{(1)}$ and $A^{(2)}$ are the same. Hence, the covariances, Σ_t, are the same for all $t \in \{1, \ldots, T\}$ and H_0 is true. If $\delta \neq 0$, the null hypothesis is not true and k_1 is the underlying true covariance change point.

In the simulation studies, we set $n = 40, 50$ and 60, with $p = 500, 750$ and 1000. The number of repeated measurements, T, was set to be 5 and 8 and set $L = 3$. The simulation results reported in Tables 2.5 and 2.6 were based on 500 simulation replications.

Table 2.5: Empirical size and power of the proposed test, percentages of simulation replications that reject the null hypothesis for data generated from a standardized Gamma distribution under the nominal level 5%

		$T = 5$			$T = 8$		
			p			p	
δ	n	500	750	1000	500	750	1000
0(size)	40	3.6	4.0	4.4	4.0	3.6	4.6
	50	4.2	5.2	4.4	5.2	4.8	4.8
	60	4.6	3.8	4.6	5.0	5.0	5.6
0.05	40	23.4	21.4	28.2	35.2	38.6	31.2
	50	38.2	36.2	33.4	47.8	50.4	47.8
	60	46.4	46.8	46.2	64.6	67.2	66.4
0.10	40	99.8	99.8	100	100	100	100
	50	100	100	100	100	100	100
	60	100	100	100	100	100	100

Table 2.5 reports the empirical size and power of the proposed test under the null and alternative hypotheses. We observe that Type I error is well controlled with the empirical sizes close to the nominal level of 5%. The results demonstrate the robustness of the proposed method for non-Gaussian distributed random vectors. When the differences between covariance matrices increase, the power of the proposed test increases accordingly. Table 2.6 reports the performance of the proposed change point identification procedure under the non-Gaussian distributed random vectors. We observe that the percentages of correct identification with non-Gaussian random vectors are similar to those under the Gaussian setup.

2.5.3 Accuracy of correlation matrix estimator of V_{nD}

This section aims to evaluate the numerical performance of the correlation matrix estimator, $\hat{V}_{nD}$, proposed immediately following Theorem 3. To measure the difference between $\hat{V}_{nD}$

Table 2.6: Percentages of correct change point identification among all rejected hypotheses for data generated from a standardized Gamma distribution

		$T=5$			$T=8$		
			p			p	
δ	n	500	750	1000	500	750	1000
	40	32.48	42.99	35.46	26.70	30.05	30.13
0.05	50	50.26	52.49	46.71	40.17	42.06	46.86
	60	49.14	55.56	57.58	48.30	50.00	52.41
	40	93.79	96.79	95.80	95.20	94.60	95.00
0.10	50	98.80	99.40	99.00	98.60	97.00	97.20
	60	99.60	99.20	99.80	99.60	99.20	99.20

and V_{nD}, we used the average component-wise quadratic distance, namely, $(T-1)^{-2}\|\hat{V}_{nD}-V_{nD}\|_F^2$. Figure 2.1 illustrates the average of $(T-1)^{-2}\|\hat{V}_{nD}-V_{nD}\|_F^2$ based on 500 simulation replications conducted in setting (I) under the null hypothesis with $T=5$. We observe that the correlation matrix estimator, $\hat{V}_{nD}$, is reliable when $n=40$. The performance further improves as the sample size increases.

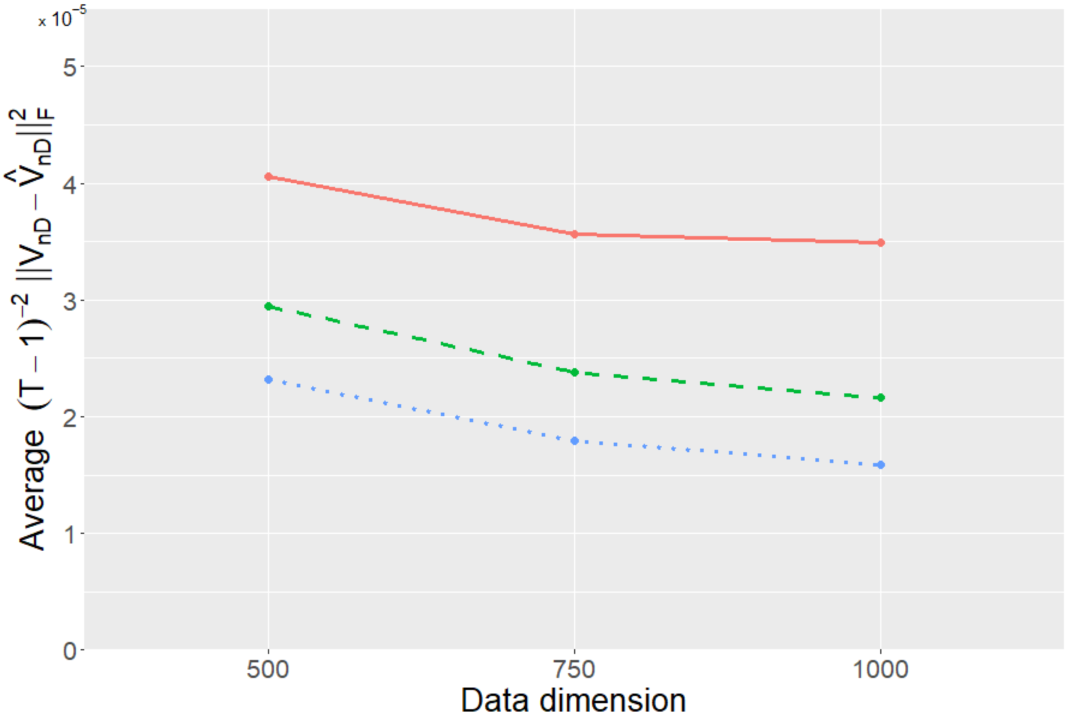

Figure 2.1: The average component-wise quadratic distance between $\hat{V}_{nD}$ and V_{nD}. The top solid line is for $n=40$; the middle dashed line is for $n=50$; the bottom dotted line is for $n=60$. The scale of the y-axis is 10^{-5}.

2.5.4 Comparison with a pair-wise based method

In this section, we compare our proposed method with a pair-wise based method that is similar to the method proposed by Zalesky et al. (2014). In the pair-wise based method, we first obtain a p-value for testing the homogeneity of each component of the covariance matrix for every pair of coordinates (u, v) with $u \leq v$ and $u, v \in \{1, \dots, p\}$, and then apply the Bonferroni correction to all the p-values to control the family-wise error rate.

In the first step, for each pair (u, v) with $u \leq v$ and $u, v \in \{1, \dots, p\}$, we test the following hypothesis

$$H_{0,uv} : \sigma_{1,uv} = \cdots = \sigma_{T,uv},$$

versus

$$H_{1,uv} : \sigma_{1,uv} = \cdots = \sigma_{k_1,uv} \neq \sigma_{k_1+1,uv} = \cdots = \sigma_{k_q,uv} \neq \sigma_{k_q+1,uv} = \cdots = \sigma_{T,uv}.$$

To test $H_{0,uv}$, we apply the statistic $\hat{D}_{nt,uv}$ defined in Section 4, and define $\hat{D}_{n,uv} = n \sum_{t=1}^{T-1} \hat{D}_{nt,uv}$. Under $H_{0,uv}$, the asymptotic distribution of $\hat{D}_{n,uv}$ is $\sum_{l=1}^{\infty} \lambda_l \chi_l^2$, where χ_l^2 are independent chi-square distributions with degree of freedom 1, and λ_l's are the eigenvalues of the kernel of $\hat{D}_{n,uv}$. In practice, one can approximate the weighted chi-square distribution using a scaled chi-square distribution. Thus, we approximate the distribution of $\hat{D}_{n,uv}$ by $b\chi_\nu^2$, where $b = \sigma_{uv}^2/(2\mu_{uv})$ and $\nu = 2\mu_{uv}^2/\sigma_{uv}^2$. Here μ_{uv} and σ_{uv}^2 are the mean and variance of $\hat{D}_{n,uv}$ under $H_{0,uv}$, respectively. The variance of $\hat{D}_{n,uv}$ under $H_{0,uv}$ is

$$\sigma_{uv}^2 = \sum_{t=1}^{T-1} \sum_{q=1}^{T-1} \sum_{s_1=1}^{t} \sum_{h_1=1}^{q} \sum_{s_2=t+1}^{T} \sum_{h_2=q+1}^{T} \sum_{\substack{k,l,\\ s,t \in \{1,2\}}} (-1)^{|k-l|+|s-t|} G_{s_k s_l h_t h_s}^{(uv)},$$

where $G_{s_k s_l h_t h_s}^{(uv)}$ is defined in Section 4. The mean of $\hat{D}_{n,uv}$ under the null $H_{0,uv}$ is

$$\mu_{uv} = \sum_{t=1}^{T-1} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{a,b \in \{1,2\}} (-1)^{|a-b|} \{C_{s_a s_b}^{(uu)} C_{s_a s_b}^{(vv)} + C_{s_a s_b}^{(uv)} C_{s_a s_b}^{(vu)}\}.$$

We then approximate the distribution of $\hat{D}_{n,uv}$ by $\hat{b}\chi_{\hat{\nu}}^2$ where $\hat{b} = \hat{\sigma}_{uv}^2/(2\hat{\mu}_{uv})$ and $\hat{\nu} = 2\hat{\mu}_{uv}^2/\hat{\sigma}_{uv}^2$. The p-value for the (u, v) pair is computed as $p_{uv} = \mathrm{pr}(\hat{b}\chi_{\hat{\nu}}^2 > \hat{D}_{n,uv})$.

In the second step, we apply the Bonferroni correction to control the family-wise error rate. Define $p_{\min} = \min_{u \leq v} p_{uv}$ as the minimum of all the pair-wise p-values. If $p_{\min} < 2\alpha/\{p(p+1)\}$, then we reject the null hypothesis on the homogeneity of covariance matrices at the α level.

To compare the proposed methods with the pair-wise based method, we conducted a simulation study using the simulation setup given in Subsection 2.5.1. The simulation results are summarized in Table 2.7. We observe that the pair-wise based method has very conservative size under the null hypothesis when sample size is less than 80, but it improves as sample size increases. Under the alternatives, the power of the pair-wise based method is low for the small sample cases, but it increases as sample size increases to 80. However, in all the cases, our proposed power-enhanced method has superior power than the pair-wise based method.

Table 2.7: Empirical size and power, percentages rejecting the null hypotheses in the simulations, for the pair-wise based test and the power-enhanced test statistic $\mathcal{M}_n^*$

		Pair-wise based test		$\mathcal{M}_n^*$	
n	p	Null	Alternative	Null	Alternative
40	500	0.2	0.2	6.0	67.6
40	750	0.0	0.4	3.6	62.8
40	1000	0.0	0.0	4.6	62.4
50	500	0.4	0.4	5.6	91.6
50	750	0.2	0.0	6.6	94.2
50	1000	0.2	0.2	3.4	97.0
60	500	0.6	12.2	4.4	98.8
60	750	0.2	4.8	5.6	99.2
60	1000	0.6	1.0	4.2	99.8
80	500	0.4	97.6	4.0	100
80	750	2.4	98.8	3.8	100
80	1000	2.0	96.8	4.6	100

2.6 An empirical study

In this section, we apply our proposed method to a time-course gene expressions data set collected by Taylor et al. (2007). The goal was to identify gene sets with significant changes in covariances over time and estimate their respective change points, should any exist. The data correspond to a study where peripheral blood mononuclear cells were collected from 69 patients with hepatitis C virus. The cells were collected once before treatment, day 0, and five times during treatment: days 1, 2, 7, 14 and 28. The treatment consisted of pegylated alpha interferon and ribavirin. More information about the experiment can be found in Taylor et al. (2007).

Prior to the application of our methodology, the data were pre-processed. The gene expressions with low quality measurements were removed if the corresponding Microarray Suite 5.0 signal transcript was classified as absent. We only kept individuals with gene expression arrays at all six time points. After pre-processing, our data set consisted of 46 individuals with gene expression arrays at days 0, 1, 2, 7, 14 and 28. The original data set can be obtained at `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7123`.

The genes were grouped into gene sets that were defined by gene ontology, which classifies genes according to attributes of the gene in three biological domains: molecular function, biological process, and cellular component (Ashburner et al., 2000). For instance, the gene ontology term labeled 0006468 is related to introducing a phosphate group onto a protein. Hence, this gene ontology term would consist of all the genes that have a role in the aforementioned biological process. A given gene can be a member of multiple gene ontologies. For example, in our processed data set, gene ontology 0006468 consists of 221 genes and gene ontology 0007155 consists of 134 genes, with 64 genes in common. After filtering the data set according to the procedure above, 159 gene ontology terms were analyzed. We applied our method to gene ontology terms with a minimum of 100 genes. Figure 2.2 illustrates the number of genes in the 159 gene ontology terms. Each gene set analyzed had a gene count much larger than the sample size of 46 patients.

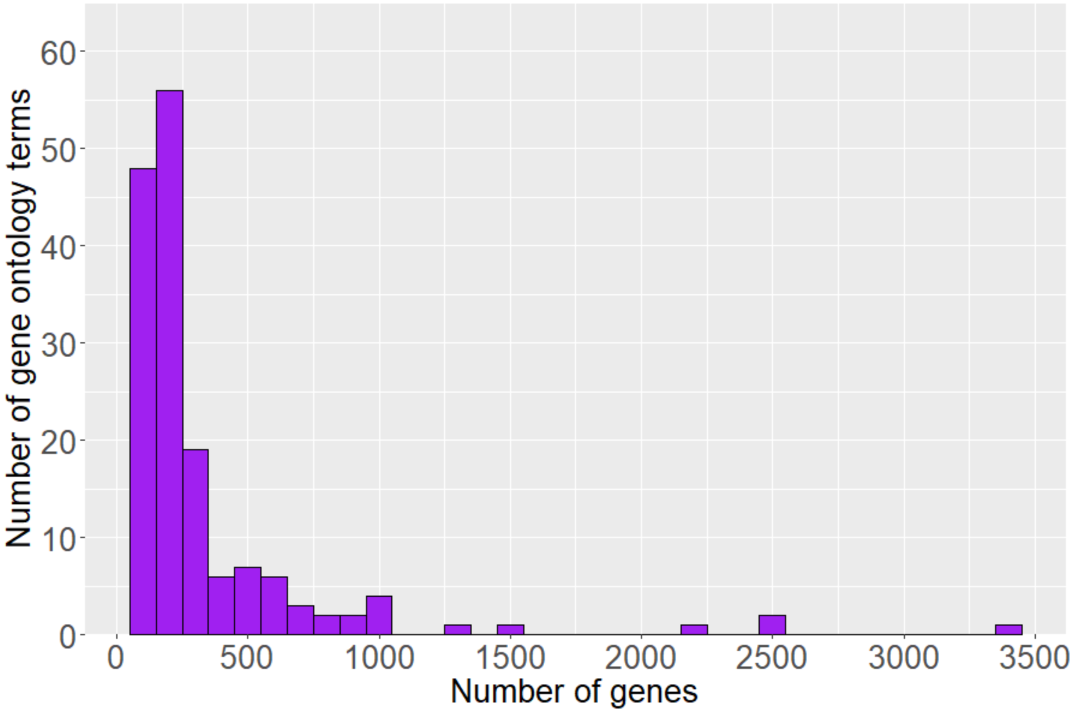

Figure 2.2: Histogram of the number of genes among the 159 gene ontology terms analyzed.

Let $Y_{it}^{(g)}$ $(i = 1, \ldots, 46; t = 1, \ldots, 6)$ be the gene expression data for the gth gene ontology term of the ith individual at time t, where $t = 1$ represents day 0, before treatment, and $t = 2, 3, 4, 5, 6$ represent the times during the treatment of hepatitis C virus with pegylated alpha interferon and ribavirin. Assume model (2.3) for each gene ontology term, $Y_{it}^{(g)} = \mu_t^{(g)} + \varepsilon_{it}^{(g)}$ for $g = 1, \ldots, 159$, where $\mu_t^{(g)}$ is an unknown mean vector and $var(\varepsilon_{it}^{(g)}) = \Sigma_t^{(g)}$. The assumptions on $\varepsilon_{it}^{(g)}$ in model (2.3) incorporate temporal dependence so that $\{\varepsilon_{it}^{(g)}\}_{t=1}^{T}$ are dependent over time. For each gene ontology term, we tested whether the covariance matrices, $\Sigma_t^{(g)}$, are the same across all t. In addition, the change points were identified for those gene ontology terms found to be significant.

For the gth gene ontology term, we computed $\hat{D}_{nt}^{(g)}/\hat{\sigma}_{(g),nt,0}$ for $t = 1, \ldots, 5$ and the covariance matrix estimation $\hat{V}_{n,D}^{(g)}$. Let $\tilde{\mathcal{M}}_n^{(g)}$ be the maximum of the standardized test statistics $\{\hat{V}_{n,D}^{(g)}\}^{-1/2}\{\hat{\sigma}_{(g),n1,0}^{-1}\hat{D}_{n1}^{(g)}, \ldots, \hat{\sigma}_{(g),n5,0}^{-1}\hat{D}_{n5}^{(g)}\}^{\mathrm{T}}$. For each gene ontology term, the local false discovery rate was estimated using $\{\tilde{\mathcal{M}}_n^{(g)}\}_{g=1}^{159}$ based on the method proposed by Efron (2007). As suggested in Efron (2007), a cutoff value of 0.20 was used for the local false discovery rate procedure. There were 10 gene ontology terms that had a local false discovery rate less than or equal to 0.20. These 10 significant gene ontology terms and their corresponding number of genes, test statistic value, estimated change points, and local false discovery rate are listed in Table 2.8. Among those gene ontology terms listed in Table 2.8, term 0008285 is associated with the reduction or stoppage of cell proliferation. This is of interest, as Kannan et al. (2011) had noted that the hepatitis C virus reduces cell proliferation. Thus, the results here suggest that treatment using pegylated alpha interferon and ribavirin has some effect on the covariances of those genes that play a role in cellular proliferation.

Table 2.8: Significant gene ontology terms, test statistic values, number of genes in each gene ontology term, identified change points and estimated local false discovery rates

GO	Number of Genes	Test Statistic Value	Change Points	Local FDR
0006511	132	11.10	4, 5	0.012
0030054	136	9.92	1, 4, 5	0.044
0042493	128	9.54	5	0.064
0008219	122	9.34	4, 5	0.076
0006357	167	9.13	1, 4	0.090
0005765	116	8.93	4	0.103
0019904	117	8.87	4, 5	0.106
0008285	148	8.75	1, 2, 5	0.115
0048471	263	8.04	1, 4, 5	0.168
0005739	661	8.04	4, 5	0.168

After identifying ten significant gene ontology terms, we applied binary segmentation to identify all change points. We discovered that eight terms have a change point at $t = 5$, day 14, eight have a change point at $t = 4$, day 7, and four terms have a change point at $t = 1$, day 0. Recall that a change point at time $t = 5$ implies the covariance matrix at time $t = 5$ is not equal to that at time $t = 6$. Hence, most of the identified changes in the covariance matrices occurred by the initial day of treatment or later in the treatment cycle. These findings complement those of Taylor et al. (2007), who observed that the majority of the genes that were altered in expression occurred at the early days of treatment and again, marginally, between treatment days 7 and 28. To illustrate the changes in covariance matrices, Figure 2.3 demonstrates the correlation networks of gene ontology term 0030054 at the six time points. We see that the correlation networks change at time points 1, 4 and 5, which is consistent with the identified change points reported in Table 2.8.

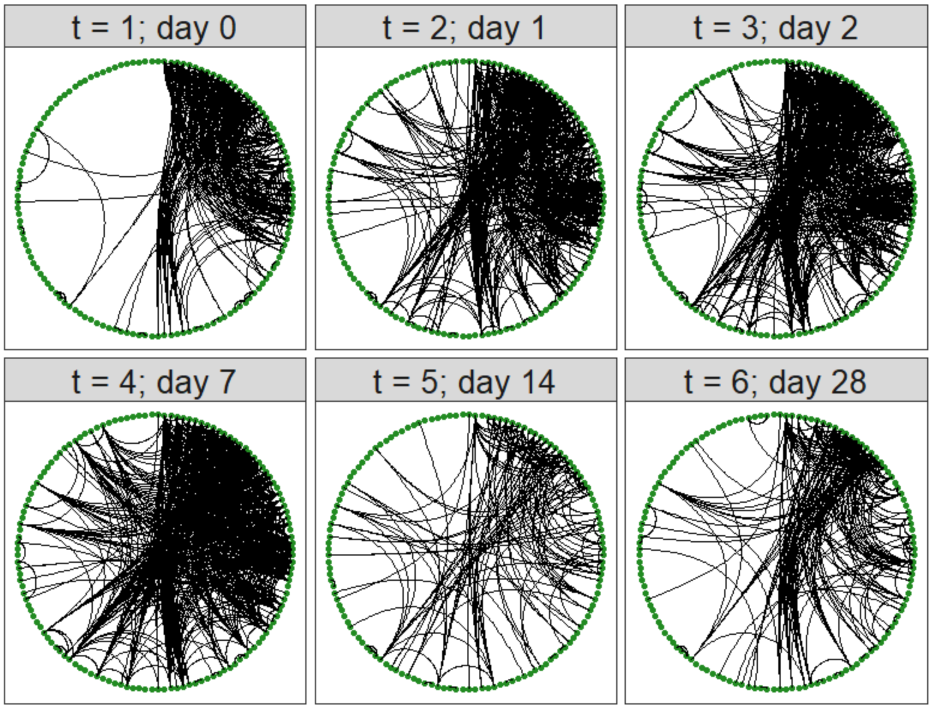

Figure 2.3: Correlation network map for gene ontology term 0030054. Each dot represents a gene within the gene ontology. A link between dots indicates a strong correlation between genes.

2.7 Technical details

2.7.1 Proofs of lemmas

In this section, we present the proofs to some lemmas used in the proofs of the main theorems. Without loss of generality, assume that $\mu_t = 0$ in our proofs for each $t \in \{1, \ldots, T\}$ because the test statistic, $\hat{D}_{nt}$, is invariant with respect to μ_t.

Lemma 1. *(i) For any symmetric matrices A and B with appropriate dimensions, we have* $\mathrm{tr}^2(AB) \leq \mathrm{tr}(A^2)\mathrm{tr}(B^2)$; *(ii) for any square matrix A,* $|\mathrm{tr}(A^2)| \leq \mathrm{tr}(AA^{\mathrm{T}})$; *and (iii) for any square matrix A,* $\|A^2\|_F^2 \leq \|A^{\mathrm{T}}A\|_F^2$ *where* $\|B\|_F^2 = \mathrm{tr}(B^{\mathrm{T}}B)$ *is the Frobenius norm of B.*

Proof. (i) Let $A = (a_{ij})$ and $B = (b_{ij})$. By the Cauchy-Schwarz inequality,

$$\mathrm{tr}(AB) = \sum_i \sum_j a_{ij} b_{ij} \leq \sum_i \Big(\sum_j a_{ij}^2 \Big)^{1/2} \Big(\sum_j b_{ij}^2 \Big)^{1/2} \leq \Big(\sum_i \sum_j a_{ij}^2 \Big)^{1/2} \Big(\sum_i \sum_j b_{ij}^2 \Big)^{1/2}.$$

Since A and B are symmetric, the right-hand side of the above inequality is the square root of $\mathrm{tr}(A^2)\mathrm{tr}(B^2)$.

(ii) Assume that $A = (a_{ij})$ is any $p \times p$ matrix. If $\mathrm{tr}(A^2) \geq 0$, because $\mathrm{tr}\{(A^{\mathrm{T}} - A)(A^{\mathrm{T}} - A)^{\mathrm{T}}\} \geq 0$ and $\mathrm{tr}\{(A^{\mathrm{T}} - A)(A^{\mathrm{T}} - A)^{\mathrm{T}}\} = 2\mathrm{tr}(A^{\mathrm{T}}A) - 2\mathrm{tr}(A^2)$, we have $|\mathrm{tr}(A^2)| \leq \mathrm{tr}(AA^{\mathrm{T}})$.

If $\mathrm{tr}(A^2) < 0$, because $\mathrm{tr}\{(A^{\mathrm{T}} + A)(A^{\mathrm{T}} + A)^{\mathrm{T}}\} \geq 0$ and $\mathrm{tr}\{(A^{\mathrm{T}} + A)(A^{\mathrm{T}} + A)^{\mathrm{T}}\} = 2\mathrm{tr}(A^{\mathrm{T}}A) + 2\mathrm{tr}(A^2) = 2\mathrm{tr}(A^{\mathrm{T}}A) - 2|\mathrm{tr}(A^2)|$, we have $|\mathrm{tr}(A^2)| \leq \mathrm{tr}(AA^{\mathrm{T}})$.

(iii) By definition, $\|A^2\|_F^2 = \mathrm{tr}(A^{\mathrm{T}}A^{\mathrm{T}}AA) = \mathrm{tr}(A^{\mathrm{T}}AAA^{\mathrm{T}})$. Since $A^{\mathrm{T}}A$ and AA^{T} are symmetric matrices, it follows by using part (i) that

$$\mathrm{tr}(A^{\mathrm{T}}AAA^{\mathrm{T}}) \leq \|A^{\mathrm{T}}A\|_F \|AA^{\mathrm{T}}\|_F = \|A^{\mathrm{T}}A\|_F^2,$$

and this completes the proof. $\square$

Lemma 2. *Define* $U_{s_1 s_2, 0} = \{n(n-1)\}^{-1} \sum_{i \neq j = 1}^{n} (Y_{is_1}^{\mathrm{T}} Y_{js_2})^2$ *for any* $s_1, s_2 \in \{1, \ldots, T\}$. *Under Condition 1, the leading order term of the covariance between* $U_{s_1 s_2, 0}$ *and* $U_{h_1 h_2, 0}$ *is* $G_n(s_1, s_2, h_1, h_2) = \mathrm{cov}(U_{s_1 s_2, 0}, U_{h_1 h_2, 0})$, *where*

$$\begin{aligned} G_n(s_1, s_2, h_1, h_2) = {} & \frac{2}{n(n-1)} \mathrm{tr}^2(C_{s_1 h_1} C_{s_2 h_2}^{\mathrm{T}}) + \frac{2}{n(n-1)} \mathrm{tr}^2(C_{s_1 h_2} C_{s_2 h_1}^{\mathrm{T}}) \\ & + \frac{2(n-2)}{n(n-1)} \sum_{u, v \in \{1,2\}} \mathrm{tr}(\Sigma_{s_{u^c}} C_{s_u h_v} \Sigma_{h_{v^c}} C_{s_u h_v}^{\mathrm{T}}). \end{aligned}$$

Denote u^c *as the complement set of* $\{u\}$. *That is,* $u^c = \{1, 2\}/\{u\}$.

Proof. Using the notation $\widetilde{\sum}$ defined in Section 2.3, we define

$$
\begin{aligned}
L_1 &= \frac{1}{(P_n^2)^2} \sum_{i \neq j = 1}^{n} E\big[\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})\}\{(Y_{ih_1}^{\mathrm{T}} Y_{jh_2})^2 - \mathrm{tr}(\Sigma_{h_1}\Sigma_{h_2})\}\big], \\
L_2 &= \frac{1}{(P_n^2)^2} \sum_{i \neq j = 1}^{n} E\big[\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})\}\{(Y_{jh_1}^{\mathrm{T}} Y_{ih_2})^2 - \mathrm{tr}(\Sigma_{h_1}\Sigma_{h_2})\}\big], \\
L_3 &= \frac{1}{(P_n^2)^2} \widetilde{\sum}_{i,j,l}\, E\big[\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})\}\{(Y_{ih_1}^{\mathrm{T}} Y_{lh_2})^2 - \mathrm{tr}(\Sigma_{h_1}\Sigma_{h_2})\}\big], \\
L_4 &= \frac{1}{(P_n^2)^2} \widetilde{\sum}_{i,j,l}\, E\big[\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})\}\{(Y_{lh_1}^{\mathrm{T}} Y_{ih_2})^2 - \mathrm{tr}(\Sigma_{h_1}\Sigma_{h_2})\}\big], \\
L_5 &= \frac{1}{(P_n^2)^2} \widetilde{\sum}_{i,j,l}\, E\big[\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})\}\{(Y_{jh_1}^{\mathrm{T}} Y_{lh_2})^2 - \mathrm{tr}(\Sigma_{h_1}\Sigma_{h_2})\}\big], \\
L_6 &= \frac{1}{(P_n^2)^2} \widetilde{\sum}_{i,j,l}\, E\big[\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})\}\{(Y_{lh_1}^{\mathrm{T}} Y_{jh_2})^2 - \mathrm{tr}(\Sigma_{h_1}\Sigma_{h_2})\}\big], \\
L_7 &= \frac{1}{(P_n^2)^2} \widetilde{\sum}_{i,j,k,l}\, E\big[\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})\}\{(Y_{kh_1}^{\mathrm{T}} Y_{lh_2})^2 - \mathrm{tr}(\Sigma_{h_1}\Sigma_{h_2})\}\big].
\end{aligned}
$$

Then $\mathrm{cov}(U_{s_1 s_2,0}, U_{h_1 h_2,0}) = L_1 + \cdots + L_7$ since $E\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2\} = \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})$. Applying standard results in multivariate analysis, we obtain

$$
\begin{aligned}
E\{(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 (Y_{ih_1}^{\mathrm{T}} Y_{jh_2})^2\} &= 2\mathrm{tr}(C_{h_1 s_1} C_{s_2 h_2} C_{h_1 s_1} C_{s_2 h_2}) + 2\mathrm{tr}(\Sigma_{s_1} C_{s_2 h_2}^{\mathrm{T}} \Sigma_{h_1} C_{s_2 h_2}) \\
&\quad + 2\mathrm{tr}^2(C_{s_2 h_2} C_{h_1 s_1}) + 2\mathrm{tr}(\Sigma_{s_2} C_{s_1 h_1} \Sigma_{h_2} C_{s_1 h_1}^{\mathrm{T}}) + \mathrm{tr}(\Sigma_{s_2}\Sigma_{s_1})\mathrm{tr}(\Sigma_{h_2}\Sigma_{h_1}).
\end{aligned}
$$

This implies that

$$
\begin{aligned}
L_1 + L_2 = \frac{2}{n(n-1)} \Big[& \mathrm{tr}\{(C_{s_2 h_2} C_{h_1 s_1})^2\} + \mathrm{tr}\{(C_{s_2 h_1} C_{h_2 s_1})^2\} + \mathrm{tr}^2(C_{s_2 h_2} C_{h_1 s_1}) \\
& + \mathrm{tr}^2(C_{s_2 h_1} C_{h_2 s_1}) + \mathrm{tr}(\Sigma_{s_1} C_{s_2 h_2} \Sigma_{h_1} C_{s_2 h_2}^{\mathrm{T}}) + \mathrm{tr}(\Sigma_{s_2} C_{s_1 h_1} \Sigma_{h_2} C_{s_1 h_1}^{\mathrm{T}}) \\
& + \mathrm{tr}(\Sigma_{s_1} C_{s_2 h_1} \Sigma_{h_2} C_{s_2 h_1}^{\mathrm{T}}) + \mathrm{tr}(\Sigma_{s_2} C_{s_1 h_2} \Sigma_{h_1} C_{s_1 h_2}^{\mathrm{T}}) \Big].
\end{aligned}
$$

Furthermore, $L_7 = 0$ and

$$
\sum_{i=3}^{6} L_i = \frac{2(n-2)}{n(n-1)} \sum_{u,v \in \{1,2\}} \mathrm{tr}(\Sigma_{s_{u^c}} C_{s_u h_v} \Sigma_{h_{v^c}} C_{s_u h_v}^{\mathrm{T}}),
$$

This with Condition 1 implies that Lemma 2 is valid. $\square$

Lemma 3. *Define* $U_{s_1s_2,1} = (1/P_n^3)\ \widetilde{\sum}_{i,j,k}\ Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1}$. *The leading term in the covariance between* $U_{s_1s_2,1}$ *and* $U_{h_1h_2,1}$ *is*

$$\begin{aligned} cov(U_{s_1s_2,1}, U_{h_1h_2,1}) = {} & \frac{4}{n^3} \sum_{u,v\in\{1,2\}} \mathrm{tr}^2(C_{s_uh_v} C_{s_{u^c}h_{v^c}}^{\mathrm{T}}) \\ & + \frac{2}{n^2} \sum_{u,v\in\{1,2\}} \mathrm{tr}(\Sigma_{s_{u^c}} C_{s_uh_v} \Sigma_{h_{v^c}} C_{s_uh_v}^{\mathrm{T}}), \end{aligned}$$

where u^c *is the complement set of* $\{u\}$. *That is,* $u^c = \{1,2\}/\{u\}$. *In addition,* $\mathrm{v}ar(\hat{D}_{nt,1}) = o\{\mathrm{v}ar(\hat{D}_{nt,0})\}$.

Proof. Because $E(U_{s_1s_2,1}) = 0$, $cov(U_{s_1s_2,1}, U_{h_1h_2,1}) = E(U_{s_1s_2,1}U_{h_1h_2,1})$. By definition,

$$\begin{aligned} U_{s_1s_2,1}U_{h_1h_2,1} = {} & \frac{1}{(P_n^3)^2} \widetilde{\sum}_{i,j,k} \widetilde{\sum}_{i_1,j_1,k_1} (Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1} + Y_{is_2}^{\mathrm{T}} Y_{js_1} Y_{js_1}^{\mathrm{T}} Y_{ks_2}) \\ & \times (Y_{i_1h_1}^{\mathrm{T}} Y_{j_1h_2} Y_{j_1h_2}^{\mathrm{T}} Y_{k_1h_1} + Y_{i_1h_2}^{\mathrm{T}} Y_{j_1h_1} Y_{j_1h_1}^{\mathrm{T}} Y_{k_1h_2}). \end{aligned}$$

According to the number of equivalent indices among two sets $\{i,j,k\}$ and $\{i_1,j_1,k_1\}$, we decompose $U_{s_1s_2,1}U_{h_1h_2,1}$ into three terms. Let $I_c = \{i,j,k\} \cup \{i_1,j_1,k_1\}$ where c represents the number of indices that are equivalent to each other in two sets $\{i,j,k\}$ and $\{i_1,j_1,k_1\}$. If there is one index equivalent,

$$\begin{aligned} I_1 = \{ & (i = i_1, j, k, j_1, k_1), (i = j_1, j, k, i_1, k_1), (i, j, k = i_1, i_1, j_1), \\ & (i, j = i_1, k, j_1, k_1), (i, j = j_1, k, i_1, k_1), (i, j = k_1, k, i_1, j_1), \\ & (i, j, k = i_1, j_1, k_1), (i, j, k = j_1, i_1, k_1), (i, j, k = k_1, i_1, j_1)\}. \end{aligned}$$

For each case within I_1, the expectation of corresponding summand in $U_{s_1s_2,1}U_{h_1h_2,1}$ is 0. If there are two indices equivalent,

$$\begin{aligned} I_2 = \{ & (i = i_1, j = j_1, k, k_1), (i = j_1, j = i_1, k, k_1), (i = i_1, k = k_1, j, j_1), \\ & (i = k_1, k = i_1, j, j_1), (j = j_1, k = k_1, i, i_1), (j = k_1, k = j_1, i, i_1), \\ & (i = i_1, j = k_1, k, j_1), (i = j_1, j = k_1, k, i_1), (i = i_1, k = j_1, j, k_1), \\ & (i = k_1, k = j_1, j, i_1), (j = j_1, k = i_1, i, k_1), (j = k_1, k = i_1, i, j_1)\}. \end{aligned}$$

Among all the cases in I_2, there exist two cases $\{(i = i_1, k = k_1, j, j_1), (i = k_1, k = i_1, j, j_1)\}$ whose expectations of the summand in $U_{s_1 s_2,1} U_{h_1 h_2,1}$ are not zero. Similarly, if there are three indices equivalent,

$$
\begin{aligned}
I_3 = \{&(i = i_1, j = j_1, k = k_1), (i = j_1, j = i_1, k = k_1), (i = k_1, j = j_1, k = i_1), \\
&(i = k_1, j = i_1, k = j_1), (i = i_1, j = k_1, k = j_1), (i = j_1, j = k_1, k = i_1)\}.
\end{aligned}
$$

Among all the cases in I_3, there are two cases $(i = i_1, j = j_1, k = k_1)$ and $(i = k_1, j = j_1, k = i_1)$ that have non-zero expectation.

In summary,

$$
\begin{aligned}
E(U_{s_1 s_2,1} U_{h_1 h_2,1}) = {}& \frac{2}{(P_n^3)^2} E\Big\{ \widetilde{\sum}_{i,k,j,j_1} (Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1} + Y_{is_2}^{\mathrm{T}} Y_{js_1} Y_{js_1}^{\mathrm{T}} Y_{ks_2}) \\
& \qquad \times (Y_{ih_1}^{\mathrm{T}} Y_{j_1 h_2} Y_{j_1 h_2}^{\mathrm{T}} Y_{kh_1} + Y_{ih_2}^{\mathrm{T}} Y_{j_1 h_1} Y_{j_1 h_1}^{\mathrm{T}} Y_{kh_2}) \Big\} \\
& + \frac{2}{(P_n^3)^2} E\Big\{ \widetilde{\sum}_{i,k,j} (Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1} + Y_{is_2}^{\mathrm{T}} Y_{js_1} Y_{js_1}^{\mathrm{T}} Y_{ks_2}) \\
& \qquad \times (Y_{ih_1}^{\mathrm{T}} Y_{jh_2} Y_{jh_2}^{\mathrm{T}} Y_{kh_1} + Y_{ih_2}^{\mathrm{T}} Y_{jh_1} Y_{jh_1}^{\mathrm{T}} Y_{kh_2}) \Big\} \\
= {}& \frac{2}{P_n^3} \sum_{u,v \in \{1,2\}} \Big[(n-3) \mathrm{tr}(\Sigma_{s_u c} C_{s_u h_v} \Sigma_{h_v c} C_{s_u h_v}^{\mathrm{T}}) + \mathrm{tr}^2(C_{s_u h_v} C_{s_u c h_v c}^{\mathrm{T}}) \\
& + \mathrm{tr}\{(C_{s_u h_v} C_{s_u c h_v c}^{\mathrm{T}})^2\} + \mathrm{tr}(\Sigma_{s_u c} C_{s_u h_v} \Sigma_{h_v c} C_{s_u h_v}^{\mathrm{T}}) \Big].
\end{aligned}
$$

This completes the proof. $\square$

Lemma 4. *Define* $U_{s_1 s_2,2} = (1/P_n^4) \widetilde{\sum}_{i,j,k,l} (Y_{is_1}^{\mathrm{T}} Y_{js_2})(Y_{ks_1}^{\mathrm{T}} Y_{ls_2})$. *For any fixed* $u, v, k, l \in \{1, 2\}$, *the covariance between* $U_{s_u s_v,2}$ *and* $U_{h_k h_l,2}$ *is*

$$
\begin{aligned}
cov(U_{s_u s_v,2}, U_{h_k h_l,2}) = {}& \frac{2}{P_n^4} \{ \mathrm{tr}^2(C_{s_u h_k} C_{s_v h_l}^{\mathrm{T}}) + \mathrm{tr}(C_{s_u h_k} C_{s_v h_l}^{\mathrm{T}} C_{s_u h_k} C_{s_v h_l}^{\mathrm{T}}) \\
& + 2\mathrm{tr}(C_{s_v h_k} C_{s_v h_l}^{\mathrm{T}} C_{s_u h_l} C_{s_u h_k}^{\mathrm{T}}) + 2\mathrm{tr}(C_{s_v h_k} C_{s_v h_l}^{\mathrm{T}} C_{s_u h_k} C_{s_u h_l}^{\mathrm{T}}) \\
& + 3\mathrm{tr}(C_{s_v h_k} C_{s_u h_l}^{\mathrm{T}} C_{s_v h_l} C_{s_u h_k}^{\mathrm{T}}) + 3\mathrm{tr}(C_{s_v h_k} C_{s_u h_l}^{\mathrm{T}}) \mathrm{tr}(C_{s_v h_l} C_{s_u h_k}^{\mathrm{T}}) \}.
\end{aligned}
$$

Moreover, $\mathrm{var}(\hat{D}_{nt,2}) = o\{\mathrm{var}(\hat{D}_{nt,0})\}$.

Proof. Because $E(U_{s_us_v,2}) = 0$, $\mathrm{cov}(U_{s_us_v,2}, U_{h_kh_l,2}) = E(U_{s_us_v,2}U_{h_kh_l,2})$. Therefore, let $\mathcal{R} = E\{(Y_{is_u}^{\mathrm{T}} Y_{js_v})(Y_{ks_u}^{\mathrm{T}} Y_{ls_v})(Y_{i_1h_k}^{\mathrm{T}} Y_{j_1h_l})(Y_{k_1h_k}^{\mathrm{T}} Y_{l_1h_l})\}$, so

$$\begin{aligned}
\mathrm{cov}(U_{s_us_v,2}, U_{h_kh_l,2}) &= \frac{1}{(P_n^4)^2} \widetilde{\sum}_{i,j,k,l} \widetilde{\sum}_{i_1,j_1,k_1,l_l} \mathcal{R} \\
&= \frac{2}{P_n^4}\{\mathrm{tr}^2(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}) + \mathrm{tr}(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}) \\
&\quad + 2\mathrm{tr}(C_{s_vh_k}C_{s_vh_l}^{\mathrm{T}}C_{s_uh_l}C_{s_uh_k}^{\mathrm{T}}) + 2\mathrm{tr}(C_{s_vh_k}C_{s_vh_l}^{\mathrm{T}}C_{s_uh_k}C_{s_uh_l}^{\mathrm{T}}) \\
&\quad + 3\mathrm{tr}(C_{s_vh_k}C_{s_uh_l}^{\mathrm{T}}C_{s_vh_l}C_{s_uh_k}^{\mathrm{T}}) + 3\mathrm{tr}(C_{s_vh_k}C_{s_uh_l}^{\mathrm{T}})\mathrm{tr}(C_{s_vh_l}C_{s_uh_k}^{\mathrm{T}})\}.
\end{aligned}$$

This completes the proof of the first part.

For the second part, write $\hat{D}_{nt,2} = w^{-1}(t)\sum_{s_1=1}^{t}\sum_{s_2=t+1}^{\mathrm{T}}\sum_{u,v\in\{1,2\}}(-1)^{|u-v|}U_{s_us_v,2}$. It follows by the first part that

$$\begin{aligned}
\mathrm{v}ar(\hat{D}_{nt,2}) = \frac{1}{w^2(t)}\frac{2}{n^4} &\sum_{s_1,s_2,h_1,h_2}^{*} \sum_{u,v,k,l\in\{1,2\}} (-1)^{|u-v|+|k-l|}\{\mathrm{tr}^2(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}) \\
&+ \mathrm{tr}(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}) + 2\mathrm{tr}(C_{s_vh_k}C_{s_vh_l}^{\mathrm{T}}C_{s_uh_l}C_{s_uh_k}^{\mathrm{T}}) \\
&+ 2\mathrm{tr}(C_{s_vh_k}C_{s_vh_l}^{\mathrm{T}}C_{s_uh_k}C_{s_uh_l}^{\mathrm{T}}) + 3\mathrm{tr}(C_{s_vh_k}C_{s_uh_l}^{\mathrm{T}}C_{s_vh_l}C_{s_uh_k}^{\mathrm{T}}) \\
&+ 3\mathrm{tr}(C_{s_vh_k}C_{s_uh_l}^{\mathrm{T}})\mathrm{tr}(C_{s_vh_l}C_{s_uh_k}^{\mathrm{T}})\}.
\end{aligned}$$

Applying the inequalities given in Lemma 1, we can show that $\mathrm{v}ar(\hat{D}_{nt,2}) = o\{\mathrm{v}ar(\hat{D}_{nt,0})\}$. This completes the proof of this Lemma. □

Lemma 5. *Let Z be an m-dimensional multivariate normally distributed random vector with mean 0 and covariance I_m. Define $M = ZZ^{\mathrm{T}} - I$. Assume A, B, C, D are matrices* *with appropriate dimensions. Then* $E\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\} = \mathrm{tr}^2(A^{\mathrm{T}}B) + \mathrm{tr}\{(A^{\mathrm{T}}B)^2\}$ *and*

$$\begin{aligned}
&\mathrm{cov}\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}}), \mathrm{tr}(CMC^{\mathrm{T}}DMD^{\mathrm{T}})\} \\
&= 2\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}(C^{\mathrm{T}}D)\mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)(C^{\mathrm{T}}D + D^{\mathrm{T}}C)\} \\
&\quad + \frac{1}{2}\mathrm{tr}^2\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)(C^{\mathrm{T}}D + D^{\mathrm{T}}C)\} + \mathrm{tr}\big[\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)(C^{\mathrm{T}}D + D^{\mathrm{T}}C)\}^2\big] \\
&\quad + 2\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)(C^{\mathrm{T}}DC^{\mathrm{T}}D + D^{\mathrm{T}}CD^{\mathrm{T}}C)\} \\
&\quad + 2\mathrm{tr}(C^{\mathrm{T}}D)\mathrm{tr}\{(C^{\mathrm{T}}D + D^{\mathrm{T}}C)(A^{\mathrm{T}}BA^{\mathrm{T}}B + B^{\mathrm{T}}AB^{\mathrm{T}}A)\} \\
&\quad + 2\mathrm{tr}\{(A^{\mathrm{T}}BA^{\mathrm{T}}B + B^{\mathrm{T}}AB^{\mathrm{T}}A)(C^{\mathrm{T}}DC^{\mathrm{T}}D + D^{\mathrm{T}}CD^{\mathrm{T}}C)\}.
\end{aligned}$$

In particular,

$$\begin{aligned}
&\mathrm{v}ar\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\} = 2\mathrm{tr}^2(A^{\mathrm{T}}B)\mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)^2\} + \frac{1}{2}\mathrm{tr}^2\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)^2\} \\
&\quad + 4\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)(A^{\mathrm{T}}BA^{\mathrm{T}}B + B^{\mathrm{T}}AB^{\mathrm{T}}A)\} \\
&\quad + 2\mathrm{tr}\{(A^{\mathrm{T}}BA^{\mathrm{T}}B + B^{\mathrm{T}}AB^{\mathrm{T}}A)^2\} + \mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)^4\}.
\end{aligned}$$

Moreover, $\mathrm{v}ar\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\} \leq K\Big[\mathrm{tr}^4(A^{\mathrm{T}}B) + \mathrm{tr}^2\{(A^{\mathrm{T}}B)^{\otimes 2}\}\Big]$ *for a constant* $K > 0$.

Proof. We first consider $E\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\}$. Because $M = ZZ^{\mathrm{T}} - I$, we have

$$\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}}) = (Z^{\mathrm{T}}A^{\mathrm{T}}BZ)^2 - Z^{\mathrm{T}}A^{\mathrm{T}}BB^{\mathrm{T}}AZ - Z^{\mathrm{T}}B^{\mathrm{T}}AA^{\mathrm{T}}BZ + \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A). \tag{2.12}$$

Taking expectation of the both sides of equation (2.12), we have

$$\begin{aligned}
E\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\} &= \mathrm{tr}^2(A^{\mathrm{T}}B) + \mathrm{tr}\{(A^{\mathrm{T}}B)^2\} + \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A) - \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A) \\
&= \mathrm{tr}^2(A^{\mathrm{T}}B) + \mathrm{tr}\{(A^{\mathrm{T}}B)^2\}.
\end{aligned}$$

Next, we consider the covariance part. Using equation (2.12), we have

$$
\begin{aligned}
&\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\mathrm{tr}(CMC^{\mathrm{T}}DMD^{\mathrm{T}}) = (Z^{\mathrm{T}}A^{\mathrm{T}}BZ)^2(Z^{\mathrm{T}}C^{\mathrm{T}}DZ)^2 \\
&- (Z^{\mathrm{T}}A^{\mathrm{T}}BZ)^2Z^{\mathrm{T}}C^{\mathrm{T}}DD^{\mathrm{T}}CZ - (Z^{\mathrm{T}}A^{\mathrm{T}}BZ)^2Z^{\mathrm{T}}D^{\mathrm{T}}CC^{\mathrm{T}}DZ \\
&- (Z^{\mathrm{T}}A^{\mathrm{T}}BZ)^2\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C) - Z^{\mathrm{T}}B^{\mathrm{T}}AA^{\mathrm{T}}BZ(Z^{\mathrm{T}}C^{\mathrm{T}}DZ)^2 \\
&+ (Z^{\mathrm{T}}B^{\mathrm{T}}AA^{\mathrm{T}}BZ)(Z^{\mathrm{T}}C^{\mathrm{T}}DD^{\mathrm{T}}CZ) + (Z^{\mathrm{T}}B^{\mathrm{T}}AA^{\mathrm{T}}BZ)(Z^{\mathrm{T}}D^{\mathrm{T}}CC^{\mathrm{T}}DZ) \\
&- (Z^{\mathrm{T}}B^{\mathrm{T}}AA^{\mathrm{T}}BZ)\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C) - Z^{\mathrm{T}}A^{\mathrm{T}}BB^{\mathrm{T}}AZ(Z^{\mathrm{T}}C^{\mathrm{T}}DZ)^2 \\
&+ Z^{\mathrm{T}}A^{\mathrm{T}}BB^{\mathrm{T}}AZ(Z^{\mathrm{T}}C^{\mathrm{T}}DD^{\mathrm{T}}CZ) + Z^{\mathrm{T}}A^{\mathrm{T}}BB^{\mathrm{T}}AZ(Z^{\mathrm{T}}D^{\mathrm{T}}CC^{\mathrm{T}}DZ) \\
&- Z^{\mathrm{T}}A^{\mathrm{T}}BB^{\mathrm{T}}AZ\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C) + \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)(Z^{\mathrm{T}}C^{\mathrm{T}}DZ)^2 \\
&- \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)(Z^{\mathrm{T}}C^{\mathrm{T}}DD^{\mathrm{T}}CZ) - \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)(Z^{\mathrm{T}}D^{\mathrm{T}}CC^{\mathrm{T}}DZ) \\
&- \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C).
\end{aligned}
$$

Define the terms in the above expression as $J_1, \ldots, J_{16}$. We consider the expectation of each J_i for $i = 1, \ldots, 16$. We have the following:

$$
\begin{aligned}
E(J_4) &= [\mathrm{tr}^2(A^{\mathrm{T}}B) + \mathrm{tr}\{(A^{\mathrm{T}}B)^2\} + \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)]\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C), \\
E(J_6) &= \mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}B)\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C) + 2\mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}BC^{\mathrm{T}}DD^{\mathrm{T}}C), \\
E(J_7) &= \mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}B)\mathrm{tr}(D^{\mathrm{T}}CC^{\mathrm{T}}D) + 2\mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}BD^{\mathrm{T}}CC^{\mathrm{T}}D), \\
E(J_8) &= E(J_{12}) = E(J_{14}) = -\mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}B)\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C), \\
E(J_{10}) &= \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C) + 2\mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}AC^{\mathrm{T}}DD^{\mathrm{T}}C), \\
E(J_{11}) &= \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)\mathrm{tr}(D^{\mathrm{T}}CC^{\mathrm{T}}D) + 2\mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}AD^{\mathrm{T}}CC^{\mathrm{T}}D), \\
E(J_{13}) &= \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)[\mathrm{tr}^2(C^{\mathrm{T}}D) + \mathrm{tr}\{(C^{\mathrm{T}}D)^2\} + \mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C)].
\end{aligned}
$$

In addition, we can show that, for any matrices A, B, C of appropriate dimensions,

$$
\begin{aligned}
E(Z^{\mathrm{T}}AZZ^{\mathrm{T}}BZZ^{\mathrm{T}}CZ) = {} & \mathrm{tr}(A)\mathrm{tr}(B)\mathrm{tr}(C) + \mathrm{tr}(A)\{\mathrm{tr}(BC) + \mathrm{tr}(B^{\mathrm{T}}C)\} \\
& + \mathrm{tr}(B)\{\mathrm{tr}(AC) + \mathrm{tr}(A^{\mathrm{T}}C)\} + \mathrm{tr}(C)\{\mathrm{tr}(AB) + \mathrm{tr}(A^{\mathrm{T}}B)\} \\
& + \mathrm{tr}\{(A + A^{\mathrm{T}})(B + B^{\mathrm{T}})(C + C^{\mathrm{T}})\}.
\end{aligned}
$$

Applying the above formula to J_2, J_3, J_5 and J_9, we obtain

$$\begin{aligned}
-E(J_2) = {} & \mathrm{tr}^2(A^{\mathrm{T}}B)\mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C) + \mathrm{tr}(A^{\mathrm{T}}B)\{\mathrm{tr}(A^{\mathrm{T}}BC^{\mathrm{T}}DD^{\mathrm{T}}C) + \mathrm{tr}(B^{\mathrm{T}}AC^{\mathrm{T}}DD^{\mathrm{T}}C)\} \\
& + \mathrm{tr}(A^{\mathrm{T}}B)\{\mathrm{tr}(A^{\mathrm{T}}BC^{\mathrm{T}}DD^{\mathrm{T}}C) + \mathrm{tr}(B^{\mathrm{T}}A^{\mathrm{T}}C^{\mathrm{T}}DD^{\mathrm{T}}C)\} \\
& + \mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C)\{\mathrm{tr}(A^{\mathrm{T}}BA^{\mathrm{T}}B) + \mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}B)\} \\
& + 2\mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)^2C^{\mathrm{T}}DD^{\mathrm{T}}C\}.
\end{aligned}$$

The expectation of J_3 is the same as $E(J_2)$ above except for changing $C^{\mathrm{T}}DD^{\mathrm{T}}C$ to $D^{\mathrm{T}}CC^{\mathrm{T}}D$. Similarly,

$$\begin{aligned}
-E(J_5) = {} & \mathrm{tr}^2(C^{\mathrm{T}}D)\mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}B) + \mathrm{tr}(C^{\mathrm{T}}D)\{\mathrm{tr}(C^{\mathrm{T}}DB^{\mathrm{T}}AA^{\mathrm{T}}B) + \mathrm{tr}(D^{\mathrm{T}}CB^{\mathrm{T}}AA^{\mathrm{T}}B)\} \\
& + \mathrm{tr}(C^{\mathrm{T}}D)\{\mathrm{tr}(C^{\mathrm{T}}DB^{\mathrm{T}}AA^{\mathrm{T}}B) + \mathrm{tr}(D^{\mathrm{T}}C^{\mathrm{T}}B^{\mathrm{T}}AA^{\mathrm{T}}B)\} \\
& + \mathrm{tr}(B^{\mathrm{T}}AA^{\mathrm{T}}B)\{\mathrm{tr}(C^{\mathrm{T}}DC^{\mathrm{T}}D) + \mathrm{tr}(D^{\mathrm{T}}CC^{\mathrm{T}}D)\} \\
& + 2\mathrm{tr}\{(C^{\mathrm{T}}D + D^{\mathrm{T}}C)^2B^{\mathrm{T}}AA^{\mathrm{T}}B\},
\end{aligned}$$

and $E(J_9)$ is the same as $E(J_5)$ with replacing $B^{\mathrm{T}}AA^{\mathrm{T}}B$ with $A^{\mathrm{T}}BB^{\mathrm{T}}A$. Finally, we can show that

$$\begin{aligned}
E(J_1) = {} & \mathrm{tr}^2(A^{\mathrm{T}}B)\mathrm{tr}^2(C^{\mathrm{T}}D) + \mathrm{tr}^2(A^{\mathrm{T}}B)[\mathrm{tr}\{(C^{\mathrm{T}}D)^2\} + \mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C)] \\
& + \mathrm{tr}^2(C^{\mathrm{T}}D)[\mathrm{tr}\{(A^{\mathrm{T}}B)^2\} + \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)] \\
& + 4\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}(C^{\mathrm{T}}D)\{\mathrm{tr}(A^{\mathrm{T}}BC^{\mathrm{T}}D) + \mathrm{tr}(B^{\mathrm{T}}AC^{\mathrm{T}}D)\} \\
& + [\mathrm{tr}\{(A^{\mathrm{T}}B)^2\} + \mathrm{tr}(A^{\mathrm{T}}BB^{\mathrm{T}}A)][\mathrm{tr}\{(C^{\mathrm{T}}D)^2\} + \mathrm{tr}(C^{\mathrm{T}}DD^{\mathrm{T}}C)] \\
& + 2\{\mathrm{tr}(A^{\mathrm{T}}BC^{\mathrm{T}}D) + \mathrm{tr}(B^{\mathrm{T}}AC^{\mathrm{T}}D)\}^2 \\
& + 2\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)(CD + D^{\mathrm{T}}C^{\mathrm{T}})^2\} \\
& + 2\mathrm{tr}(C^{\mathrm{T}}D)\mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)^2(CD + D^{\mathrm{T}}C^{\mathrm{T}})\} \\
& + \mathrm{tr}\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)^2(CD + D^{\mathrm{T}}C^{\mathrm{T}})^2\} + \mathrm{tr}[\{(A^{\mathrm{T}}B + B^{\mathrm{T}}A)(CD + D^{\mathrm{T}}C^{\mathrm{T}})\}^2].
\end{aligned}$$

Summarizing the above $E(J_i)$'s, we obtain $E\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\mathrm{tr}(CMC^{\mathrm{T}}DMD^{\mathrm{T}})\}$. From this result and (2.12), we can obtain the covariance between $\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})$ and

$\mathrm{tr}(CMC^{\mathrm{T}}DMD^{\mathrm{T}})$. The variance is a special case of the covariance. This completes the first part of the Lemma.

Next, we prove the inequality given in the second part. Using the Cauchy-Schwarz inequality and Lemma 1,

$$2\mathrm{tr}^2(A^{\mathrm{T}}B)\mathrm{tr}\{(A^{\mathrm{T}}B+B^{\mathrm{T}}A)^2\} \le K\mathrm{tr}^2(A^{\mathrm{T}}B)\mathrm{tr}\{(A^{\mathrm{T}}B)^{\otimes 2}\}$$

and

$$\begin{aligned}
&2\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}\{(A^{\mathrm{T}}B+B^{\mathrm{T}}A)(A^{\mathrm{T}}BA^{\mathrm{T}}B+B^{\mathrm{T}}AB^{\mathrm{T}}A)\}\\
&\le 2\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}^{1/2}\{(A^{\mathrm{T}}B+B^{\mathrm{T}}A)^2\}\mathrm{tr}^{1/2}\{(A^{\mathrm{T}}BA^{\mathrm{T}}B+B^{\mathrm{T}}AB^{\mathrm{T}}A)^2\}\\
&\le K\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}^{1/2}\{(A^{\mathrm{T}}B)^{\otimes 2}\}\mathrm{tr}^{1/2}\{(A^{\mathrm{T}}BA^{\mathrm{T}}B)^{\otimes 2}\}\\
&\le K\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}^{1/2}\{(A^{\mathrm{T}}B)^{\otimes 2}\}\mathrm{tr}\big[\{(A^{\mathrm{T}}B)^{\mathrm{T}}(A^{\mathrm{T}}B)\}^2\big]\\
&\le K\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}^{3/2}\{(A^{\mathrm{T}}B)^{\otimes 2}\}.
\end{aligned}$$

Moreover, $\mathrm{tr}\{(A^{\mathrm{T}}B+B^{\mathrm{T}}A)^4\} \le \mathrm{tr}^2\{(A^{\mathrm{T}}B+B^{\mathrm{T}}A)^2\} \le K\mathrm{tr}^2\{(A^{\mathrm{T}}B)^{\otimes 2}\}$. In summary,

$$\begin{aligned}
\mathrm{v}ar\{\mathrm{tr}(AMA^{\mathrm{T}}BMB^{\mathrm{T}})\} &\le K\Big[\mathrm{tr}(A^{\mathrm{T}}B)\mathrm{tr}^{1/2}\{(A^{\mathrm{T}}B)^{\otimes 2}\}+\mathrm{tr}\{(A^{\mathrm{T}}B)^{\otimes 2}\}\Big]^2\\
&\le K\Big[\mathrm{tr}^4(A^{\mathrm{T}}B)+\mathrm{tr}^2\{(A^{\mathrm{T}}B)^{\otimes 2}\}\Big].
\end{aligned}$$

This finishes the proof of this Lemma. $\square$

Define

$$\begin{aligned}
V_{n0}(s_1,s_2,h_1,h_2) &= \frac{4}{n(n-1)}\sum_{u,v,k,l\in\{1,2\}}(-1)^{-|u-v|-|k-l|}\mathrm{tr}^2(C_{s_uh_k}C^{\mathrm{T}}_{s_vh_l}),\\
V_{n1}(s_1,s_2,h_1,h_2) &= \frac{8(n-2)}{n(n-1)}\sum_{u,v\in\{1,2\}}(-1)^{|u-v|}\mathrm{tr}\{(\Sigma_{s_1}-\Sigma_{s_2})C_{s_uh_v}(\Sigma_{h_1}-\Sigma_{h_2})C^{\mathrm{T}}_{s_uh_v}\}.
\end{aligned}$$

Lemma 6. *Let* $W_{s_1s_2}=U_{s_1s_1,0}+U_{s_2s_2,0}-U_{s_1s_2,0}-U_{s_2s_1,0}$. *The covariance between* $W_{s_1s_2}$ *and* $W_{h_1h_2}$ *is* $V_u(s_1,s_2,h_1,h_2)$, *where* $V_u(s_1,s_2,h_1,h_2)=V_{n0}(s_1,s_2,h_1,h_2)$ $+V_{n1}(s_1,s_2,h_1,h_2)$ *and* $V_{n0}(s_1,s_2,h_1,h_2)$ *is the covariance between* $W_{s_1s_2}$ *and* $W_{h_1h_2}$ *under* H_0.

Proof. Let $G_n(\cdot)$ be the function defined in Lemma 2. It then follows,

$$V_u(s_1, s_2, h_1, h_2) = \sum_{u,v,k,l\in\{1,2\}} (-1)^{-|u-v|-|k-l|} G_n(s_u, s_v, h_k, h_l).$$

Applying Lemma 2, we have

$$\begin{aligned} V_u(s_1, s_2, h_1, h_2) = \frac{2}{n(n-1)} &\sum_{u,v,k,l\in\{1,2\}} (-1)^{-|u-v|-|k-l|} \big\{\mathrm{tr}^2(C_{s_u h_k} C^{\mathrm{T}}_{s_v h_l}) \\ + \mathrm{tr}^2(C_{s_u h_l} C^{\mathrm{T}}_{s_v h_k})\big\} + \frac{2(n-2)}{n(n-1)} &\sum_{u,v,k,l\in\{1,2\}} (-1)^{-|u-v|-|k-l|} \big\{\mathrm{tr}(\Sigma_{s_u} C_{s_v h_l} \Sigma_{h_k} C^{\mathrm{T}}_{s_v h_l}) \\ + \mathrm{tr}(\Sigma_{s_v} C_{s_u h_k} \Sigma_{h_l} C^{\mathrm{T}}_{s_u h_k}) + \mathrm{tr}(\Sigma_{s_u} C_{s_v h_k} &\Sigma_{h_l} C^{\mathrm{T}}_{s_v h_k}) + \mathrm{tr}(\Sigma_{s_v} C_{s_u h_l} \Sigma_{h_k} C^{\mathrm{T}}_{s_u h_l})\big\}. \end{aligned}$$

Hence,

$$\begin{aligned} V_u(s_1, s_2, h_1, h_2) = \frac{4}{n(n-1)} &\sum_{u,v,k,l\in\{1,2\}} (-1)^{-|u-v|-|k-l|} \mathrm{tr}^2(C_{s_u h_k} C^{\mathrm{T}}_{s_v h_l}) \\ + \frac{8(n-2)}{n(n-1)} &\sum_{u,v,k,l\in\{1,2\}} (-1)^{-|u-v|-|k-l|} \mathrm{tr}(\Sigma_{s_u} C_{s_v h_l} \Sigma_{h_k} C^{\mathrm{T}}_{s_v h_l}). \end{aligned}$$

After some algebra, one can show the second term in the above expression is equivalent to $V_{n1}(s, h, h_1, h_2)$.

Under H_0, $V_{n1}(s, h, h_1, h_2) = 0$. Therefore, $V_u(s, h, h_1, h_2) = V_0(s, h, h_1, h_2)$ is the covariance under H_0. This completes the proof of Lemma 6. □

2.7.2 Proofs of main results

In this section, we present proofs for the main results of Chapter 2. By definition, $\hat{D}_{nt}$ can be expressed as $\hat{D}_{nt} = \hat{D}_{nt,0} - 2\hat{D}_{nt,1} + \hat{D}_{nt,2}$, where for $k = 0, 1$ and 2,

$$\hat{D}_{nt,k} = \frac{1}{t(T-t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} (U_{s_1 s_1,k} + U_{s_2 s_2,k} - U_{s_1 s_2,k} - U_{s_2 s_1,k}). \qquad (2.13)$$

Here $U_{s_1 s_2,k}$ was defined in Section 2.3.

Proof of Theorem 1. Based on the definition of $\hat{D}_{nt}$, the expectation of $\hat{D}_{nt}$ is

$$E(\hat{D}_{nt}) = \frac{1}{t}\sum_{s=1}^{t} \mathrm{tr}(\Sigma_s^2) + \frac{1}{T-t}\sum_{h=t+1}^{T} \mathrm{tr}(\Sigma_h^2) - \frac{2}{t(T-t)}\sum_{s=1}^{t}\sum_{h=t+1}^{T} \mathrm{tr}(\Sigma_s\Sigma_h) = D_t.$$

We next calculate the order of the variance of $\hat{D}_{nt}$. By using the definition of $\hat{D}_{nt}$, write $\hat{D}_{nt}$ as $\hat{D}_{nt} = \hat{D}_{nt,0} - 2\hat{D}_{nt,1} + \hat{D}_{nt,2}$. By Lemmas 3 and 4, it follows that $\hat{D}_{nt,1} = o_p(\hat{D}_{nt,0})$ and $\hat{D}_{nt,2} = o_p(\hat{D}_{nt,0})$. Therefore, it suffices to compute the variance of $\hat{D}_{nt,0}$. Using Lemma 6,

$$\sigma_{nt}^2 = w^{-2}(t) \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \mathrm{cov}(W_{s_1s_2}, W_{h_1h_2}) = w^{-2}(t) \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} V_u(s_1,s_2,h_1,h_2).$$

This completes the proof of Theorem 1. $\square$

Proof of Theorem 2. By Theorem 1, it is sufficient to establish the asymptotic normality of $\hat{D}_{nt,0}$. We first write $\hat{D}_{nt,0}$ into a martingale. Define $A_{js_u} = Y_{js_u}Y_{js_u}^{\mathrm{T}} - \Sigma_{s_u}$, and

$$G_{nj} = \frac{1}{t(T-t)}\sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\sum_{u,v\in\{1,2\}} (-1)^{|u-v|}\{Y_{is_u}^{\mathrm{T}}A_{js_v}Y_{is_u} - \mathrm{tr}(\Sigma_{s_u}A_{js_v})\},$$

$$Q_{ni} = \frac{1}{t(T-t)}\sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\sum_{u,v\in\{1,2\}} (-1)^{|u-v|}\{Y_{is_u}^{\mathrm{T}}\Sigma_{s_v}Y_{is_u} - \mathrm{tr}(\Sigma_{s_u}\Sigma_{s_v})\}.$$

Let $Z_{ni} = Z_{ni}^{(1)} + Z_{ni}^{(2)}$, where $Z_{ni}^{(1)} = 2\sum_{j=1}^{i-1} G_{nj}/\{n(n-1)\}$ and $Z_{ni}^{(2)} = 4Q_{ni}/n$. Then, $\hat{D}_{nt,0} - D_t = \sum_{i=1}^{n} Z_{ni}$,

Let $\mathscr{F}_k$ be the σ-algebra generated by $\sigma\{Y_1,\ldots,Y_k\}$ where $Y_i = \{Y_{i1},\ldots,Y_{iT}\}$ is the collection of Y for the i-th sample. It follows that $E(Z_{nk}|\mathscr{F}_{k-1}) = 0$. Therefore, Z_{nk} is a sequence of martingale difference with respect to $\mathscr{F}_k$.

Let $\sigma_{ni}^2 = E(Z_{ni}^2|\mathscr{F}_{i-1})$. To prove the asymptotic normality, we check two following conditions (Hall and Hedye, 1980):

Condition (a) $\sum_{i=1}^{n} \sigma_{ni}^2/\mathrm{var}(\hat{D}_{nt}) \xrightarrow{p} 1$;

Condition (b) $\sum_{i=1}^{n} E(Z_{ni}^4)/\mathrm{var}^2(\hat{D}_{nt}) \to 0$.

We first prove Condition (a). Consider $E(\sum_{i=1}^n \sigma_{ni}^2) = \sum_{i=1}^n E(Z_{ni}^2) = \sum_{i=1}^n \mathrm{var}(Z_{ni})$. Furthermore, $\mathrm{var}(\hat{D}_{nt}) = \sum_{i=1}^n E(Z_{ni}^2) + 2E\{\sum_{i<j} Z_{ni}E(Z_{nj}|\mathscr{F}_{j-1})\} = \sum_{i=1}^n E(Z_{ni}^2)$. Thus, we have $E(\sum_{i=1}^n \sigma_{ni}^2) = \mathrm{var}(\hat{D}_{nt})$. It suffices to show $\mathrm{var}(\sum_{i=1}^n \sigma_{ni}^2) = o\{\mathrm{var}^2(\hat{D}_{nt})\}$.

Now we obtain σ_{ni}^2 as

$$\sigma_{ni}^2 = E(Z_{ni}^2|\mathscr{F}_{i-1}) = \binom{n}{2}^{-2} \sum_{j=1}^{i-1}\sum_{j_1=1}^{i-1} E(G_{nj}G_{nj_1}|\mathscr{F}_{i-1}) + \frac{4}{n^2}E(Q_{ni}^2|\mathscr{F}_{i-1})$$
$$+ \frac{1}{n}\frac{4}{\binom{n}{2}}E\big(Q_{ni}\sum_{j=1}^{i-1} G_{nj}|\mathscr{F}_{i-1}\big) = R_{ni,1} + R_{ni,2} + R_{ni,3}.$$

Recall that $A_{js_1} = Y_{js_1}Y_{js_1}^{\mathrm{T}} - \Sigma_{s_1}$. We can further show that $R_{ni,2}$ is a constant and has no impact on $\mathrm{var}(\sum_{i=1}^n \sigma_{ni}^2)$. Moreover,

$$R_{ni,1} = \frac{2}{\binom{n}{2}^2 w^2(t)} \sum_{j=1}^{i-1}\sum_{j_1=1}^{i-1} \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{u,v,k,l\in\{1,2\}} (-1)^{|u-v|+|k-l|}\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}})$$
$$= R_{ni,1}^{(0)} + R_{ni,1}^{(1)},$$

where $R_{ni,1}^{(0)}$ corresponds to summation of the terms where $j = j_1$ and $R_{ni,1}^{(1)}$ is the summation of the terms where $j \neq j_1$. To prove Condition (a), it suffices to show that

(a1) $\mathrm{var}(\sum_{i=1}^n R_{ni,1}^{(0)}) = o(\sigma_{nt}^4)$,

(a2) $\mathrm{var}(\sum_{i=1}^n R_{ni,1}^{(1)}) = o(\sigma_{nt}^4)$ and

(a3) $\mathrm{var}(\sum_{i=1}^n R_{ni,3}) = o(\sigma_{nt}^4)$.

We first show (a1). We have

$$\mathrm{var}\Big(\sum_{i=1}^n R_{ni,1}^{(0)}\Big) \le \frac{C}{w^4(t)n^5}\mathrm{var}\Big\{ \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{\substack{u,v,\\ k,l\in\{1,2\}}} (-1)^{|u-v|+|k-l|}\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{jh_l}C_{s_uh_k}^{\mathrm{T}})\Big\}$$
$$\le Cw^{-4}(t)n^{-5} \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{\substack{u,v,\\ k,l\in\{1,2\}}} \mathrm{var}\Big\{\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{jh_l}C_{s_uh_k}^{\mathrm{T}})\Big\}.$$

Applying Lemma 5 and using the fact

$$\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{jh_l}C_{s_uh_k}^{\mathrm{T}}) = \mathrm{tr}\{\Gamma_{s_u}^{\mathrm{T}}\Gamma_{s_v}(Z_jZ_j^{\mathrm{T}} - I)\Gamma_{s_v}^{\mathrm{T}}\Gamma_{s_u}\Gamma_{h_k}^{\mathrm{T}}\Gamma_{h_l}(Z_jZ_j^{\mathrm{T}} - I)\Gamma_{h_l}^{\mathrm{T}}\Gamma_{h_k}\},$$

we have

$$\mathrm{v}ar\{\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{jh_l}C_{s_uh_k}^{\mathrm{T}})\} \le C\big[\mathrm{tr}^4(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}) + \mathrm{tr}^2\{(\Gamma_{s_v}^{\mathrm{T}}C_{s_uh_k}\Gamma_{h_l})^{\otimes 2}\}\big].$$

Under Condition 1, $\mathrm{var}(\sum_{i=1}^n R_{ni,1}^{(0)}) = o(\sigma_{nt}^4)$. This completes the proof of (a1).

We next show (a2). Because of $j \ne j_1$, $E\{\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}})\} = 0$. It follows that

$$\begin{aligned}\mathrm{var}(\sum_{i=1}^n R_{ni,1}^{(1)}) &\le \frac{C}{w^4(t)n^4}\mathrm{var}\Big\{\sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}}(-1)^{|u-v|+|k-l|}\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}})\Big\}\\ &\le Cn^{-4}w^{-4}(t)\sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}}\mathrm{var}\big\{\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}})\big\}.\end{aligned}$$

Similar to Lemma 5, we obtain

$$\begin{aligned}&\mathrm{var}\{\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}})\} = 2\mathrm{tr}^2\{(\Gamma_{s_v}^{\mathrm{T}}C_{s_uh_k}\Gamma_{h_l})^{\otimes 2}\} + 6\mathrm{tr}\{(\Gamma_{s_v}^{\mathrm{T}}C_{s_uh_k}\Gamma_{h_l})^{\otimes 4}\}\\ &\quad - 4\mathrm{tr}\{(\Gamma_{s_v}^{\mathrm{T}}C_{s_uh_k}\Gamma_{h_l}\Gamma_{s_v}^{\mathrm{T}}C_{s_uh_k}\Gamma_{h_l})^{\otimes 2}\} \le C\mathrm{tr}^2\{(\Gamma_{s_v}^{\mathrm{T}}C_{s_uh_k}\Gamma_{h_l})^{\otimes 2}\},\end{aligned}$$

where $Q^{\otimes 2} = QQ^{\mathrm{T}}$ and $Q^{\otimes 4} = QQ^{\mathrm{T}}QQ^{\mathrm{T}}$. By Condition 1, we have $\mathrm{var}(\sum_{i=1}^n R_{ni,1}^{(1)}) = o(\sigma_{nt}^4)$, which proves (a2).

For proving (a3), a direct computation shows that

$$\begin{aligned}\sum_{i=1}^n R_{ni,3} &= \frac{16}{n\binom{n}{2}}\sum_{j=1}^n\frac{(n-j)}{w^2(t)}\sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}}(-1)^{|u-v|+|k-l|}\mathrm{tr}(C_{s_uh_k}\Sigma_{h_l}C_{s_uh_k}^{\mathrm{T}}A_{js_v})\\ &= \frac{16}{n\binom{n}{2}}\sum_{j=1}^n\frac{(n-j)}{w^2(t)}\sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}}\sum_{u,k\in\{1,2\}}(-1)^{|u-k|}\mathrm{tr}\{(A_{js_1} - A_{js_2})C_{s_uh_k}(\Sigma_{h_1} - \Sigma_{h_2})C_{s_uh_k}^{\mathrm{T}}\}.\end{aligned}$$

Then,

$$\begin{aligned}\operatorname{var}\Big(\sum_{i=1}^{n} R_{ni,3}\Big) &\le \frac{C}{n^3 w^4(t)} \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{u,k\in\{1,2\}} E\big[\operatorname{tr}^2\{(A_{js_1}-A_{js_2})C_{suh_k}(\Sigma_{h_1}-\Sigma_{h_2})C_{suh_k}^{\mathrm{T}}\}\big] \\ &= \frac{C}{n^3 w^4(t)} \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{u,k\in\{1,2\}} \operatorname{tr}\Big\{\Big(\sum_{v=1}^{2}(-1)^v \Gamma_{sv}^{\mathrm{T}} C_{suh_k}(\Sigma_{h_1}-\Sigma_{h_2})C_{suh_k}^{\mathrm{T}}\Gamma_{sv}\Big)^2\Big\}.\end{aligned}$$

Therefore, by Condition 2, $\operatorname{var}(\sum_{i=1}^{n} R_{ni}) = o(\sigma_{nt}^4)$. Thus, Condition (a) is valid.

To check Condition (b), we compute

$$E(Z_{ni}^4) = E\{E(Z_{ni}^4|\mathscr{F}_{i-1})\} \le Cn^{-8}E\Big[E\Big\{\Big(\sum_{j=1}^{i-1} G_{nj}\Big)^4|\mathscr{F}_{i-1}\Big\}\Big] + Cn^{-4}E(Q_{ni}^4) = J_{1i}+J_{2i},$$

where $E\{(\sum_{j=1}^{i-1} G_{nj})^4|\mathscr{F}_{i-1}\} = \sum_{j=1}^{i-1} E(G_{nj}^4|\mathscr{F}_{i-1}) + \sum_{j\neq j_1}^{i-1} E(G_{nj}^2 G_{nj_1}^2|\mathscr{F}_{i-1})$. Using the definition of G_{nj}, one obtain

$$G_{nj}^2 = \frac{1}{w^2(t)} \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{u,v,k,l\in\{1,2\}} (-1)^{|u-v|+|k-l|}\operatorname{tr}(A_{is_u}A_{js_v})\operatorname{tr}(A_{ih_k}A_{jh_l}).$$

For any two symmetric matrices A and B, $E[\{Z^{\mathrm{T}}AZ-\operatorname{tr}(A)\}^2\{Z^{\mathrm{T}}BZ-\operatorname{tr}(B)\}^2] = 4\{\operatorname{tr}(A^2)\operatorname{tr}(B^2)+2\operatorname{tr}^2(AB)\} + 16\{2\operatorname{tr}(A^2B^2)+\operatorname{tr}(ABAB)\}$ and $\operatorname{tr}[\{Z^{\mathrm{T}}AZ-\operatorname{tr}(A)\}^4] \le C\operatorname{tr}^2(A^2)$. Accordingly,

$$\begin{aligned}J_{1i} &\le \frac{C}{w^4(t)n^8}\sum_{j=1}^{i-1}\sum_{s_1,s_2}^{*}\sum_{u,v\in\{1,2\}} E\{\operatorname{tr}^4(A_{is_u}A_{js_v})\} \\ &\quad + \frac{C}{w^4(t)n^8}\sum_{j\neq j_1}^{i-1}\sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}} E\{\operatorname{tr}^2(A_{is_u}A_{js_v})\operatorname{tr}^2(A_{ih_k}A_{j_1h_l})\} \le \sum_{k=1}^{5} J_{1i}^{(k)},\end{aligned}$$

where

$$
\begin{aligned}
J_{1i}^{(1)} =& \frac{C}{w^4(t)n^8}\sum_{j=1}^{i-1}\sum_{s_1,s_2}^{*}\sum_{u,v\in\{1,2\}} E\big\{\mathrm{tr}^2(\Sigma_{s_u}A_{js_v}\Sigma_{s_u}A_{js_v})\big\},\\
J_{1i}^{(2)} =& \frac{C}{w^4(t)n^8}\sum_{j\neq j_1}^{i-1}\sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}} E\big\{\mathrm{tr}(A_{js_v}\Sigma_{s_u}A_{js_v}\Sigma_{s_u})\mathrm{tr}(A_{j_1h_l}\Sigma_{h_k}A_{j_1h_l}\Sigma_{h_k})\big\},\\
J_{1i}^{(3)} =& \frac{C}{w^4(t)n^8}\sum_{j\neq j_1}^{i-1}\sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}} E\big\{\mathrm{tr}^2(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}})\big\},\\
J_{1i}^{(4)} =& \frac{C}{w^4(t)n^8}\sum_{j\neq j_1}^{i-1}\sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}} E\big\{\mathrm{tr}(\Gamma_{s_u}^{\mathrm{T}}A_{js_v}\Gamma_{s_u}\Gamma_{s_u}^{\mathrm{T}}A_{js_v}\times\\
&\qquad\qquad\qquad\qquad\qquad \Gamma_{s_u}\Gamma_{h_k}^{\mathrm{T}}A_{j_1h_l}\Gamma_{h_k}\Gamma_{h_k}^{\mathrm{T}}A_{j_1h_l}\Gamma_{h_k})\big\},\\
J_{1i}^{(5)} =& \frac{C}{w^4(t)n^8}\sum_{j\neq j_1}^{i-1}\sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}} E\big\{\mathrm{tr}(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}}A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}})\big\}.
\end{aligned}
$$

Consider the first term, $J_{1i}^{(1)}$, in the above inequality. By Lemma 5,

$$
\begin{aligned}
var\{\mathrm{tr}(A_{js_v}\Sigma_{s_u}A_{js_v}\Sigma_{s_u})\} \le{}& C\Big[\mathrm{tr}^4(\Sigma_{s_u}\Sigma_{s_v}) + 4\mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v}\Sigma_{s_u}\Sigma_{s_v})\\
&+\mathrm{tr}\{(\Gamma_{s_v}^{\mathrm{T}}\Sigma_{s_u}\Gamma_{s_v})^4\} + \mathrm{tr}\{(\Gamma_{s_v}^{\mathrm{T}}\Sigma_{s_u}\Gamma_{s_v}\Gamma_{s_v}^{\mathrm{T}}\Sigma_{s_u}\Gamma_{s_v})^{\otimes 2}\}\Big],
\end{aligned}
$$

and $E\Big[\mathrm{tr}(\Sigma_{s_u}A_{js_v}\Sigma_{s_u}A_{js_v})\Big] = \mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v}) + \mathrm{tr}(\Sigma_{s_u}\Sigma_{s_v}\Sigma_{s_u}\Sigma_{s_v})$. Thus,

$$
\begin{aligned}
E\{\mathrm{tr}^2(A_{js_v}\Sigma_{s_u}A_{js_v}\Sigma_{s_u})\} &\le C\Big[\mathrm{tr}^4(\Sigma_{s_u}\Sigma_{s_v}) + \mathrm{tr}^2\{(\Sigma_{s_u}\Sigma_{s_v})^2\} + \mathrm{tr}\{(\Sigma_{s_v}\Sigma_{s_u})^4\}\Big]\\
&\le C\mathrm{tr}^4(\Sigma_{s_u}\Sigma_{s_v}).
\end{aligned}
$$

Therefore,

$$
\sum_{i=1}^{n} J_{1i}^{(1)} \le \frac{C}{w^4(t)n^8}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{s_1,s_2}^{*}\sum_{u,v\in\{1,2\}}\mathrm{tr}^4(\Sigma_{s_u}\Sigma_{s_v}) = o(\sigma_{nt}^4). \tag{2.14}
$$

For the second and third terms, $J_{1i}^{(2)}$ and $J_{1i}^{(3)}$, consider the following inequality:

$$
\mathrm{tr}^2(A_{js_v}C_{s_uh_k}A_{j_1h_l}C_{s_uh_k}^{\mathrm{T}}) \le \mathrm{tr}(A_{js_v}\Sigma_{s_u}A_{js_v}\Sigma_{s_u})\mathrm{tr}(A_{j_1h_l}\Sigma_{h_k}A_{j_1h_l}\Sigma_{h_k}).
$$

Thus, we have

$$\begin{aligned}\sum_{i=1}^{n}(J_{1i}^{(2)} + J_{1i}^{(3)}) &\leq \frac{C}{w^4(t)n^8}\sum_{i=1}^{n}\sum_{j\neq j_1}^{i-1}\Big[\sum_{s_1,s_2}^{*}\sum_{u,v\in\{1,2\}}\{\mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v}) + \mathrm{tr}\{(\Sigma_{s_u}\Sigma_{s_v})^2\}\Big]^2\\ &\leq \frac{C}{w^4(t)n^8}\sum_{i=1}^{n}\sum_{j\neq j_1}^{i-1}\Big\{\sum_{s_1,s_2}^{*}\sum_{u,v\in\{1,2\}}\mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v})\Big\}^2 = o(\sigma_{nt}^4). \qquad (2.15)\end{aligned}$$

Next, we consider the fourth term $J_{1i}^{(4)}$. We first note the following results

$$\begin{aligned}&E[\mathrm{tr}\{A(Z_1Z_1^{\mathrm{T}} - I)A^{\mathrm{T}}C(Z_1Z_1^{\mathrm{T}} - I)C^{\mathrm{T}}D(Z_2Z_2^{\mathrm{T}} - I)D^{\mathrm{T}}F(Z_2Z_2 - I)F^{\mathrm{T}}\}]\\ &= \mathrm{tr}(A^{\mathrm{T}}C)\mathrm{tr}(D^{\mathrm{T}}F)\mathrm{tr}(DF^{\mathrm{T}}AC^{\mathrm{T}}) + \mathrm{tr}(A^{\mathrm{T}}C)\mathrm{tr}(FD^{\mathrm{T}}FD^{\mathrm{T}}AC^{\mathrm{T}})\\ &\quad + \mathrm{tr}(D^{\mathrm{T}}F)\mathrm{tr}(CA^{\mathrm{T}}CA^{\mathrm{T}}DF^{\mathrm{T}}) + \mathrm{tr}(FD^{\mathrm{T}}FD^{\mathrm{T}}CA^{\mathrm{T}}CA^{\mathrm{T}})\end{aligned}$$

for matrices A, C, D and F with appropriate dimensions. Then we can obtain

$$\begin{aligned}&E\big\{\mathrm{tr}(\Gamma_{s_u}^{\mathrm{T}}A_{js_v}\Gamma_{s_u}\Gamma_{s_u}^{\mathrm{T}}A_{js_v}\Gamma_{s_u}\Gamma_{h_k}^{\mathrm{T}}A_{j_1h_l}\Gamma_{h_k}\Gamma_{h_k}^{\mathrm{T}}A_{j_1h_l}\Gamma_{h_k})\big\}\\ &= \mathrm{tr}(\Sigma_{s_v}\Sigma_{s_u})\{\mathrm{tr}(\Sigma_{h_l}\Sigma_{h_k})\mathrm{tr}(C_{s_uh_k}\Sigma_{h_l}C_{s_uh_k}^{\mathrm{T}}\Sigma_{s_v}) + \mathrm{tr}(\Sigma_{h_l}\Sigma_{h_k}\Sigma_{h_l}C_{s_uh_k}^{\mathrm{T}}\Sigma_{s_v}C_{s_uh_k})\}\\ &+ \mathrm{tr}(\Sigma_{h_l}\Sigma_{h_k})\mathrm{tr}(\Sigma_{s_v}\Sigma_{s_u}\Sigma_{s_v}C_{s_uh_k}\Sigma_{h_l}C_{s_uh_k}^{\mathrm{T}}) + \mathrm{tr}(\Sigma_{h_l}\Sigma_{h_k}\Sigma_{h_l}C_{s_uh_k}^{\mathrm{T}}\Sigma_{s_v}\Sigma_{s_u}\Sigma_{s_v}C_{s_uh_k})\\ &\leq \mathrm{tr}^{3/2}(\Sigma_{s_v}\Sigma_{s_u})\mathrm{tr}^{3/2}(\Sigma_{h_l}\Sigma_{h_k}) + \mathrm{tr}^{3/2}(\Sigma_{s_v}\Sigma_{s_u})\mathrm{tr}^{1/2}\{(\Sigma_{h_k}\Sigma_{h_l})^4\}\\ &+ \mathrm{tr}^{3/2}(\Sigma_{h_l}\Sigma_{h_k})\mathrm{tr}^{1/2}\{(\Sigma_{s_u}\Sigma_{s_v})^4\} + \mathrm{tr}^{1/2}\{(\Sigma_{h_k}\Sigma_{h_l})^4\}\mathrm{tr}^{1/2}\{(\Sigma_{s_u}\Sigma_{s_v})^4\}.\end{aligned}$$

It follows that

$$\sum_{i=1}^{n}J_{1i}^{(4)} \leq \frac{C}{w^4(t)n^8}\sum_{i=1}^{n}\sum_{j\neq j_1}^{i-1}\Big\{\sum_{s_1,s_2}^{*}\sum_{u,v\in\{1,2\}}\mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v})\Big\}^2 = o(\sigma_{nt}^4). \qquad (2.16)$$

By Lemma 5,

$$
\begin{aligned}
E\big[\mathrm{tr}\{(C^{\mathrm{T}}_{s_u h_k} A_{js_v} C_{s_u h_k} A_{j_1 h_l})^2\}\big] &= E\big[\mathrm{tr}^2(\Gamma^{\mathrm{T}}_{s_v} C_{s_u h_k} A_{j_1 h_l} C^{\mathrm{T}}_{s_u h_k} \Gamma_{s_v}) \\
&\quad + \mathrm{tr}\{(\Gamma^{\mathrm{T}}_{s_v} C_{s_u h_k} A_{j_1 h_l} C^{\mathrm{T}}_{s_u h_k} \Gamma_{s_v})^2\}\big] \\
&= \{Z_j^{\mathrm{T}} \Gamma^{\mathrm{T}}_{h_l} C^{\mathrm{T}}_{s_u h_k} \Sigma_{s_v} C_{s_u h_k} \Gamma_{h_l} Z_j \\
&\quad - \mathrm{tr}(\Sigma_{h_l} C^{\mathrm{T}}_{s_u h_k} \Sigma_{s_v} C_{s_u h_k})\}^2 \\
&\quad + \mathrm{tr}(C_{s_u h_k} A_{j_1 h_l} C^{\mathrm{T}}_{s_u h_k} \Sigma_{s_v} C_{s_u h_k} A_{j_1 h_l} C^{\mathrm{T}}_{s_u h_k} \Sigma_{s_v}) \\
&= 3\mathrm{tr}\{(\Gamma^{\mathrm{T}}_{h_l} C^{\mathrm{T}}_{s_u h_k} \Gamma_{s_v})^{\otimes 4}\} + \mathrm{tr}^2\{(\Gamma^{\mathrm{T}}_{h_l} C^{\mathrm{T}}_{s_u h_k} \Gamma_{s_v})^{\otimes 2}\}.
\end{aligned}
$$

This along with Condition 1 together implies $\sum_{k=1}^{n} J_{1k}^{(5)} = o(\sigma_{nt}^4)$, and further with equations (2.14) and (2.15) implies $\sum_{k=1}^{n} J_{1k} = o(\sigma_{nt}^4)$.

Finally, we consider J_{2i}. We write Q_{ni} as

$$
Q_{ni} = \frac{1}{w(t)} \sum_{s_1, s_2}^{*} \Big[Z_i^{\mathrm{T}} \Big\{ \sum_{u=1}^{2} (-1)^{u-1} \Gamma^{\mathrm{T}}_{s_u} (\Sigma_{s_1} - \Sigma_{s_2}) \Gamma_{s_u} \Big\} Z_i^{\mathrm{T}} - \mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2})(\Sigma_{s_1} + \Sigma_{s_2})\} \Big],
$$

where $\sum_{u=1}^{2} (-1)^{u-1} \Gamma^{\mathrm{T}}_{s_u} (\Sigma_{s_1} - \Sigma_{s_2}) \Gamma_{s_u} = (\Gamma_{s_1} + \Gamma_{s_2})^{\mathrm{T}} (\Sigma_{s_1} - \Sigma_{s_2})(\Gamma_{s_1} - \Gamma_{s_2})$. Using Proposition A.1 in Chen et al. (2010),

$$
J_{2i} \leq \frac{C}{n^4 w^2(t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{\mathrm{T}} \mathrm{tr}^2 \Big[\{(\Gamma_{s_1} + \Gamma_{s_2})^{\mathrm{T}} (\Sigma_{s_1} - \Sigma_{s_2})(\Gamma_{s_1} - \Gamma_{s_2})\}^{\otimes 2} \Big].
$$

As a result, $\sum_{i=1}^{n} J_{2i} = o(\sigma_{nt}^4)$. Condition (b) is valid. This completes the proof of the asymptotic normality of $\hat{D}_{nt}$. $\square$

Proof of Theorem 3. Using the continuous mapping theorem, we only need to prove the joint multivariate normality of $\{\hat{D}_{nt}\}_{t=1}^{T-1}$. Let $a = (a_1, \ldots, a_{T-1})^{\mathrm{T}}$ be any non-zero constant vector of length $T-1$. By the Cramér-Wold device, it suffices to show that $\sum_{t=1}^{T-1} a_t \hat{D}_{nt}$ is asymptotically normal under H_0. By Lemma 6, the variance of $\sum_{t=1}^{T-1} a_t \hat{D}_{nt}$ is $\sigma_{0T}^2 = \mathrm{var}(\sum_{t=1}^{T-1} a_t \hat{D}_{nt}) = \sum_{t=1}^{T-1} \sum_{q=1}^{T-1} a_t a_q Q_{n,tq}$. Then we wish to show $\sigma_{0T}^{-1} \sum_{t=1}^{T-1} a_t \hat{D}_{nt} \to N(0,1)$ in distribution. The asymptotic normality of $\sum_{t=1}^{T-1} a_t \hat{D}_{nt}$ can be shown by using the martingale central limit theory, which is very similar to the proof of Theorem 2. Therefore, we omit the details. $\square$

Proof of Theorem 4. Assume that the alternative H_1^* is true. First, for $t \in \{1, \dots k_1 - 1\}$,

$$
\begin{aligned}
E(\hat{D}_{nt}) &= \mathrm{tr}(\Sigma_1^2) + \frac{1}{T-t}\sum_{h=t+1}^{k_1}\mathrm{tr}(\Sigma_h^2) + \frac{1}{T-t}\sum_{h=k_1+1}^{T}\mathrm{tr}(\Sigma_h^2) \\
&\quad - \frac{2}{t(T-t)}\sum_{s=1}^{t}\Big\{\sum_{h=t+1}^{k_1}\mathrm{tr}(\Sigma_s\Sigma_h) + \sum_{h=k_1+1}^{T}\mathrm{tr}(\Sigma_s\Sigma_h)\Big\} \\
&= \Big\{1 + \frac{k_1 - t}{T-t} - \frac{2t(k_1 - t)}{t(T-t)}\Big\}\mathrm{tr}(\Sigma_1^2) + \frac{T-k_1}{T-t}\mathrm{tr}(\Sigma_T^2) - \frac{2t(T-k_1)}{t(T-t)}\mathrm{tr}(\Sigma_1\Sigma_T) \\
&= \frac{T-k_1}{T-t}\mathrm{tr}\big\{(\Sigma_1 - \Sigma_T)^2\big\}.
\end{aligned}
$$

Similarly, if $k_1 \le t$, then $E(\hat{D}_{nt}) = k_1\mathrm{tr}\big\{(\Sigma_1 - \Sigma_T)^2\big\}/t$.

Define $B(C) = \{t \in \{1, \dots, T-1\} : |t - k_1| \ge C\beta_n/(n\Delta_n)\}$ for some constant C. To establish the rate of convergence of the change point estimator $\hat{k}_1$, we need to show, for any $\epsilon > 0$, there exist a constant C such that $\mathrm{pr}\{|\hat{k}_1 - k_1| > C\beta_n/(n\Delta_n)\} < \epsilon$. This is equivalent to show that $\mathrm{pr}\{\hat{k}_1 \in B(C)\} < \epsilon$. Since the event $\{\hat{k}_1 \in B(C)\} \subset \{\max_{t\in B(C)} \hat{D}_{nt} > \hat{D}_{nk_1}\}$, $\mathrm{pr}\{\hat{k}_1 \in B(C)\} \le \mathrm{pr}\{\max_{t\in B(C)} \hat{D}_{nt} > \hat{D}_{k_1}\}$. Thus, it suffices to show, for any $\epsilon > 0$, there exist a constant C such that

$$
\mathrm{pr}\{\max_{t\in B(C)} \hat{D}_{nt} > \hat{D}_{k_1}\} \le \sum_{t\in B(C)} \mathrm{pr}(\hat{D}_{nt} - D_{k_1} > \hat{D}_{nk_1} - D_{k_1}) < \epsilon. \tag{2.17}
$$

Under H_1^*, we have

$$
\begin{aligned}
\hat{D}_{nt} - D_{k_1} = \hat{D}_{nt} - D_t + D_t - D_{k_1} &= \hat{D}_{nt} - D_t + \{r(t; k_1) - 1\}\mathrm{tr}\{(\Sigma_1 - \Sigma_T)^2\} \\
&= \hat{D}_{nt} - D_t - |t - k_1|G(t; k_1)\mathrm{tr}\{(\Sigma_1 - \Sigma_T)^2\},
\end{aligned}
$$

where $G(t;k_1) = \{1/(T-t)\}I(1 \leq t \leq k_1) + (1/t)I(k_1+1 \leq t \leq T-1)$. Then, for $t \in B(C)$,

$$\begin{aligned}
\mathrm{pr}(\hat{D}_{nt} > \hat{D}_{nk_1}) &\leq \mathrm{pr}\{|\hat{D}_{nt} - D_t| > |t-k_1|G(t;k_1)\Delta_n/2\} \\
&\quad + \mathrm{pr}\{|\hat{D}_{nk_1} - D_{k_1}| > |t-k_1|G(t;k_1)\Delta_n/2\} \\
&\leq \mathrm{pr}\{|\sigma_{nt}^{-1}(\hat{D}_{nt} - D_t)| > C\beta_n G(t;k_1)/\sqrt{(4V_{0t} + 8nV_{1t})}\} \\
&\quad + \mathrm{pr}\{|\sigma_{nk_1}^{-1}(\hat{D}_{nk_1} - D_{k_1})| > C\beta_n G(t;k_1)/\sqrt{(4V_{0k_1} + 8nV_{1k_1})}\}.
\end{aligned}$$

For any t and some constant C_1, $\beta_n > C_1\sqrt{(4V_{0t} + 8nV_{1t})}$. Furthermore, $w(t)$ and $G(t;k_1)$ are bounded away from zero for $t \in B(C)$. Thus, by Chebyshev's inequality,

$$\mathrm{pr}(\hat{D}_{nt} > \hat{D}_{nk_1}) \leq \mathrm{pr}\{|\sigma_{nt}^{-1}(\hat{D}_{nt} - D_t)| > C\} + \mathrm{pr}\{|\sigma_{nk_1}^{-1}(\hat{D}_{nk_1} - D_{k_1})| > C\} \leq \frac{2}{C^2} < \frac{\epsilon}{T},$$

for large enough C. Therefore, (2.17) is true. This finishes the proof of Theorem 4. □

Proof of Theorem 5. Let $k_0 = 0$ and $k_{q+1} = T$. Denote the common covariances between the change points k_j and k_{j+1} as $\tilde{\Sigma}_j$ for $j = 0, \ldots, q$. To show that $\max_t D_t$ is at one of the change points, it is enough to show that $\max_t D_t$ cannot be attained at any time points except change points $k_1, \ldots, k_q$. Thus, we need to show that the maximum of D_t is not attainable for t in the following sets: (1) $t \in \{1, \ldots, k_1 - 1\}$; (2) $t \in \{k_q + 1, \ldots, T-1\}$; and (3) $t \in \{k_l + 1, \ldots, k_{l+1} - 1\}$ for some $l \in \{1, \ldots, q-1\}$. We do not need to consider case (1) if $k_1 = 1$ or case (3) if $k_q = T-1$. Without loss of generality, we assume $k_1 > 1$ and $k_q < T-1$ in the following proof.

First, if $t \in \{1, \ldots, k_1 - 1\}$, then using the definition of D_t, we have

$$\begin{aligned}
D_t &= \frac{1}{t(T-t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{k_1} \|\Sigma_{s_1} - \Sigma_{s_2}\|_F^2 + \frac{1}{t(T-t)} \sum_{s_1=1}^{t} \sum_{s_2=k_1+1}^{T} \|\Sigma_{s_1} - \Sigma_{s_2}\|_F^2 \\
&= \frac{1}{T-t} \sum_{s_2=k_1+1}^{T} \|\tilde{\Sigma}_0 - \Sigma_{s_2}\|_F^2
\end{aligned}$$

which is an increasing function of t in this scenario. Therefore, the maximum value of D_t will not be at any $t \in \{1, \ldots, k_1 - 1\}$.

Second, if $t \in \{k_q + 1, \ldots, T-1\}$, then

$$
\begin{aligned}
D_t &= \frac{1}{t(T-t)} \sum_{s_1=1}^{k_q} \sum_{s_2=t+1}^{T} \|\Sigma_{s_1} - \Sigma_{s_2}\|_F^2 + \frac{1}{t(T-t)} \sum_{s_1=k_q+1}^{t} \sum_{s_2=t+1}^{T} \|\Sigma_{s_1} - \Sigma_{s_2}\|_F^2 \\
&= \frac{1}{t} \sum_{s_1=1}^{k_q} \|\tilde{\Sigma}_q - \Sigma_{s_1}\|_F^2
\end{aligned}
$$

which is a decreasing function of t. Therefore, the maximum value of D_t will not be at any $t \in \{k_q + 1, \ldots, T-1\}$.

At last, let us consider the third case with $t \in \{k_l + 1, \ldots, k_{l+1} - 1\}$ for some $l \in \{1, \ldots, q-1\}$. We rewrite D_t as

$$
\begin{aligned}
D_t = \frac{1}{t(T-t)} \Big\{ & \sum_{i=0}^{l-1} \sum_{j=l+1}^{q} (k_{i+1} - k_i)(k_{j+1} - k_j) \|\tilde{\Sigma}_i - \tilde{\Sigma}_j\|_F^2 \\
& + (t - k_l) \sum_{j=l+1}^{q} (k_{j+1} - k_j) \|\tilde{\Sigma}_l - \tilde{\Sigma}_j\|_F^2 + (k_{l+1} - t) \sum_{i=0}^{l-1} (k_{i+1} - k_i) \|\tilde{\Sigma}_i - \tilde{\Sigma}_l\|_F^2 \Big\}.
\end{aligned}
$$

Since $\|\tilde{\Sigma}_i - \tilde{\Sigma}_j\|_F^2 = \|\tilde{\Sigma}_i - \tilde{\Sigma}_l\|_F^2 + \|\tilde{\Sigma}_l - \tilde{\Sigma}_j\|_F^2 + 2\mathrm{tr}\{(\tilde{\Sigma}_i - \tilde{\Sigma}_l)(\tilde{\Sigma}_l - \tilde{\Sigma}_j)\}$, we further write D_t as

$$
D_t = \frac{1}{t(T-t)} \big\{ 2\Delta + tA + (T-t)B \big\},
$$

where

$$
\Delta = \sum_{i=0}^{l-1} \sum_{j=l+1}^{q} (k_{i+1} - k_i)(k_{j+1} - k_j) \mathrm{tr}\{(\tilde{\Sigma}_i - \tilde{\Sigma}_l)(\tilde{\Sigma}_l - \tilde{\Sigma}_j)\},
$$

$A = \sum_{j=l+1}^{q} (k_{j+1} - k_j) \|\tilde{\Sigma}_l - \tilde{\Sigma}_j\|_F^2$ and $B = \sum_{i=0}^{l-1} (k_{i+1} - k_i) \|\tilde{\Sigma}_i - \tilde{\Sigma}_l\|_F^2$. Then we can use the fact that $1/\{t(T-t)\} = (1/T)\{1/t + 1/(T-t)\}$ to further write D_t as

$$
D_t = \frac{1}{t}\Big(A + \frac{2\Delta}{T}\Big) + \frac{1}{T-t}\Big(B + \frac{2\Delta}{T}\Big).
$$

We will consider four cases, (a)-(d), according to the signs of $A + 2\Delta/T$ and $B + 2\Delta/T$.

(a) If $A + 2\Delta/T \geq 0$ and $B + 2\Delta/T \leq 0$, then D_t is a decreasing function of t. In this case, the maximum of D_t will not be at any t for $t \in \{k_l + 1, \ldots, k_{l+1} - 1\}$.

(b) If $A + 2\Delta/T \leq 0$ and $B + 2\Delta/T \geq 0$, then D_t is an increasing function of t. In this case, the maximum of D_t will not be at any t for $t \in \{k_l + 1, \ldots, k_{l+1} - 1\}$.

(c) If $A + 2\Delta/T > 0$ and $B + 2\Delta/T > 0$, then the derivative of D_t with respect to t is

$$D'_t = \frac{1}{t^2(T-t)^2}\{(B-A)t^2 + 2(A + \frac{2\Delta}{T})Tt - (A + \frac{2\Delta}{T})T^2\}.$$

The denominator of D'_t is always positive for $t \in \{k_l + 1, \ldots, k_{l+1} - 1\}$. Thus, to determine the sign of D'_t, we only need to know the sign of the numerator of D'_t.

The numerator of D'_t is a quadratic form of t. To know the sign of the numerator of D'_t, we consider two cases: (i) $B > A$ and (ii) $B < A$. In the case (i) with $B > A$, one of the solution of $t^2(T-t)^2 D'_t = 0$ is less than 0, another solution t_0 is greater than 0. If $t_0 \in (k_l, k_{l+1})$, then D'_t is negative for $k_l < t < t_0$ and positive for $t_0 < t < k_{l+1}$. This implies that the function D_t decreases for $k_l < t < t_0$ and increases for $t_0 < t < k_{l+1}$. Therefore, D_t attains its minimum at t_0 and the maximum of D_t will not be attained within (k_l, k_{l+1}). If $t_0 \notin (k_l, k_{l+1})$, then D'_t is either always negative or always positive for $t \in (k_l, k_{l+1})$. In this case, D_t is a monotonic function of t and hence the maximum of D_t will not be attained within (k_l, k_{l+1}).

In the case (ii) with $B < A$, it can be shown that $t^2(T-t)^2 D'_t = 0$ has two solutions, $t_1, t_2 = T\big[(A + 2\Delta/T)/(A - B) \pm \sqrt{\{(A + 2\Delta/T)/(A - B) - 1/2\}^2 - 1/4}\big]$. Here, t_1, t_2 corresponds to the positive and negative sign, respectively. Because $B + 2\Delta/T > 0$, $(A + 2\Delta/T)/(A - B) > 1$. It follows that $t_2 > T$. Similar to the case of $B > A$, if $t_1 \in (k_l, k_{l+1})$, the function D_t decreases for $k_l < t < t_1$ and increases for $t_1 < t < k_{l+1}$. Therefore, D_t attains its minimum at t_0 and the maximum of D_t will not be attained within (k_l, k_{l+1}). If $t_1 \notin (k_l, k_{l+1})$, D_t is a monotone function of t and hence the maximum of D_t will not be attained within (k_l, k_{l+1}). In summary, the maximum of D_t will not be attained within (k_l, k_{l+1}) if $A + 2\Delta/T > 0$ and $B + 2\Delta/T > 0$.

(d) If $A + 2\Delta/T < 0$ and $B + 2\Delta/T < 0$, then $2\Delta/T < 0$ because $A > 0$ and $B > 0$.

Thus, $A - 2|\Delta|/T < 0$. Using the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
A < 2|\Delta|/T &\le \sum_{i=0}^{l-1} \sum_{j=l+1}^{q} (k_{i+1} - k_i)(k_{j+1} - k_j)(\|\tilde{\Sigma}_i - \tilde{\Sigma}_l\|_F^2 + \|\tilde{\Sigma}_l - \tilde{\Sigma}_j\|_F^2)/T \\
&= \{(T - k_{l+1})A + k_l B\}/T.
\end{aligned}
$$

The above inequality implies that $A/B < k_l/k_{l+1} < 1$. On the other hand, $B < 2|\Delta|/T \le \{(T - k_{l+1})A + k_l B\}/T$, which implies that $A/B > (T - k_l)/(T - k_{l+1}) > 1$. This is a contradiction. Therefore, case (d) is not possible.

By the results of (a)-(d), the maximum of D_t will not attain within $\{k_l + 1, \ldots, k_{l+1} - 1\}$ for case (3). Thus, the proof is completed. $\square$

Proof of Theorem 6. At the beginning of the binary segmentation algorithm, we have $\mathcal{M}_n[1,T] > W_{\alpha_n}[1,T]$ with probability one because, for any $t \in \{1, \ldots, T-1\}$,

$$
\begin{aligned}
&\text{pr}(\mathcal{M}_n[1,T] > W_{\alpha_n}[1,T]) \ge \text{pr}(\sigma_{nt,0}^{-1}[1,T]\hat{D}_{nt}[1,T] > W_{\alpha_n}[1,T]) \\
&= \text{pr}\big\{\sigma_{nt}^{-1}[1,T](\hat{D}_{nt}[1,T] - D_t[1,T]) > \sigma_{nt}^{-1}[1,T](\sigma_{nt,0}[1,T]W_{\alpha_n}[1,T] - D_t[1,T])\big\} \\
&= 1 - \Phi\big\{\sigma_{nt}^{-1}[1,T](\sigma_{nt,0}[1,T]W_{\alpha_n}[1,T] - D_t[1,T])\big\} \to 1,
\end{aligned}
$$

where we used the condition $W_{\alpha_n} = o(\mathrm{m}SNR)$ in Theorem 6. Therefore, using Theorems 4 and 5, one change point in $\{k_1, \ldots, k_q\}$ will be detected and estimated with probability 1 because $\beta_n[1,T] = o(nD_{k_s}[1,T])$ for some $s \in \{1, \ldots, q\}$. Each subsequence satisfies the condition $W_{\alpha_n} = o(\mathrm{m}SNR)$ in Theorem 6 and hence the detection continues.

Suppose we have detected less than q change points. By the assumptions in this theorem, there exists a segment, $\{l_1 + 1, \ldots, l_2\}$, that contains a change point, k_s, such that $W_{\alpha_n} = o(\mathrm{m}SNR)$ and $\beta_n[(l_1+1), l_2] = o\{nD_{k_s}[(l_1+1), l_2]\}$ hold. Therefore, by similar arguments as above, a change point will be detected and estimated consistently in the segment. Thus, $\hat{q} \ge q$. Once $\hat{q}$ reaches q, all subsequent segments have end points at the change points and two boundary points $1, k_1, \ldots, k_q, T$. Then, by Theorem 3, $\mathcal{M}_n[l_1, l_2] < W_{\alpha_n}$ with probability one as $\alpha_n \to 0$. This implies that no additional change point will be detected. The proof is completed. $\square$

CHAPTER 3

COVARIANCE CHANGE POINT DETECTION AND IDENTIFICATION WITH HIGH-DIMENSIONAL FUNCTIONAL DATA

3.1 Introduction

Access to high-dimensional data has exploded in recent years due to technological improvements and cost reductions. High throughput technology has facilitated the collection of genomics data, with more variables being measured than ever before. In addition, the reductions in cost have allowed measurements to be taken over time, as is the case in time-course microarray studies. Similarly, functional neuroimaging studies repeatedly measure a massive number of variables throughout the duration of a medical experiment. Time-course microarray data and functional neuroimaging data are just two examples of applications that beget high-dimensional longitudinal, or functional, data, where a large number of variables are repeatedly measured on a small number of experimental units. Throughout this chapter we focus on high-dimensional dense functional data, where the number of repeated measurements is large (Ramsay 1982).

Functional magnetic resonance image (fMRI) data is an important example of high-dimensional functional data. In a task-based fMRI study, individuals perform various tasks while the fMRI machine records blood-oxygen-level dependent (BOLD) signals throughout their brain. These tasks may be passive or active. For example, subjects may be shown a movie, a sequence of pictures, or asked to respond to questions. In contrast, a resting-state fMRI does not involve any subject engagement, but aims to investigate the brain's functional organization through the BOLD signal measurements. In the course of an fMRI study, the human brain is partitioned into small uniform cubes, also known as voxels, that are about the size of 1–3mm^3. For each voxel, a BOLD measurement is recorded at each time point. A cluster of voxels is known as a node or region of interest, where clusters can be defined for anatomical region of interest analysis or spherical region of interest analysis. BOLD signal measurements are repeatedly recorded for each of about 100,000 brain voxels between 100 to 2000 times for a single subject. The number of repeated measurements typically depends on the fMRI scanner and duration of the task-based or resting-state experiment. To enable population inference, multiple subjects are included in an fMRI study. Rather than analyze all 100,000 voxels, doctors may be interested in specific anatomical regions of the brain. However, a region of interest will still have voxel BOLD signal measurements at the order of 100. In addition to the sheer size of the data, fMRI data exhibit complex spatiotemporal dependence. For a given subject, BOLD measurements in neighboring voxels are correlated, as are BOLD measurements for a given voxel but across time points. The high-dimensional and dependent structure make statistical modeling, testing, and analysis a challenge.

One major interest in neuroscience is to understand functional connectivity or dynamic functional connectivity at an individual or group level across time points (Kundu et al. 2018). We refer to dynamic functional connectivity as the changing relationships between spatially separated brain regions across experimental time points. In particular, we are interested in studying dynamic functional connectivity across individuals. Traditional functional connectivity assumes stationary relationships between nodes throughout the experiment. To characterize the functional connectivity at a given time point, the covariance matrix, or precision matrix, of BOLD signals serve as a proxy for the within and between brain node neural activity. As a result, dynamic functional connectivity of the brain can be explored via a procedure that assesses covariance matrix stationarity.

The purpose of this chapter is to develop a robust statistical procedure to detect and identify change points among covariance matrices in high-dimensional functional data. Assume $Y_{it} = (Y_{it1}, \ldots, Y_{itp})^{\mathrm{T}}$ is a p-dimensional random vector with mean vector μ_t and covariance matrix Σ_t. In the context of an fMRI study, Y_{it} $(i = 1, \ldots, n;\ t = 1, \ldots, T)$ represents the p BOLD signal measurements for the ith individual at the tth time point, where p, T, and n are typically at the order of 100,000, 100, and 10, respectively. For a specific region of interest in the brain or for region of interest network analysis, p may be at the order of 100. Our proposed procedure aims to answer two questions. First, does a temporal change exist among covariance matrices? This corresponds to a covariance change point detection problem that can be posed in the form of a statistical hypothesis test

$$\begin{aligned} &H_0 : \Sigma_1 = \cdots = \Sigma_T \qquad \text{versus} \\ &H_1 : \Sigma_1 = \cdots = \Sigma_{\tau_1} \neq \Sigma_{\tau_1+1} = \cdots = \Sigma_{\tau_q} \neq \Sigma_{\tau_q+1} = \cdots = \Sigma_T, \end{aligned} \tag{3.1}$$

where $\tau_k < T$ ($k = 1, \ldots, q < \infty$) are the unknown change point locations. Second, if a temporal change does exist, can we determine its location and the locations of all possible changes? This suggests a change point identification problem that aims to estimate the unknown locations of τ_ks. Although we consider a high-dimensional setting, we do not require a sparsity assumption for Σ_t, and we allow the complex spatiotemporal dependence present in high-dimensional functional data. In the context of fMRI studies, our proposed procedure will first determine if functional connectivity is stationary. If not, our change point identification procedure will partition the functional data into stationary sequences with regards to the covariance matrices.

Testing covariance matrices is a classical problem in multivariate statistical analysis. Muirhead (2005) and Anderson (2003) detailed multivariate tests for covariance matrices, including testing the homogeneity of several covariance matrices. However, these tests rely on likelihood ratios, and they require the sample size to exceed the number of variables measured. Recent work done by Schott (2007), Srivastava and Yanagihara (2010), and Li and Chen (2012) addressed the lack of an appropriate testing procedure for covariance matrices in a high-dimensional setting. More recently, Ahmad (2017) and Zhang et al. (2018) generalized aspects of the aforementioned works to an independent multi-sample test for high-dimensional covariance matrices. All of the research in testing high-dimensional covariance matrices since Schott's 2007 pioneering procedure have addressed the high-dimensional challenges. However, none have focused on how to incorporate temporal dependence in a high-dimensional setting. Therefore, none of the previously mentioned methods are applicable to high-dimensional functional data.

Researchers in neuroscience have developed a few methods to study the dynamic functional brain connectivity for single patients and populations. However, in general, their methods are ad hoc and lack the theoretical rigor to ensure a robust inference procedure. Some neuroscience approaches were detailed in Chapter 2. Most of the existing work studies dynamic functional connectivity for an individual. For example, Monti et al. (2014) developed a sliding window approach based on pair-wise correlations to study the dynamic functional connectivity. Their approach was based off a single subject and is not directly applicable to study the common dynamic functional connectivity for a population. Kundu et al. (2018) developed a procedure to test (3.1) with the aim of studying group level brain dynamic functional connectivity in a task-based fMRI experiment. To detect and identify change points, Kundu et al. (2018) first compute all pair-wise correlations between p nodes at each time point. Thus, at each time point they obtain $p(p-1)/2$ sample pair-wise correlations that they stack as a vector. Next, they apply a generalized fused Lasso (Tibshirani et al. 2005) approach to the multivariate time series of sample correlations. The fused Lasso was developed for an ordered set of covariates, and as is the case with Lasso, it also involves a penalty parameter. To tune the penalty parameter they use a lowess fit, which also depends on a smoothing parameter. Based on the fused Lasso, the number of change points is a function of the penalty parameter's value. A small value leads to more identified change points, whereas a large value leads to a fewer number of identified change points. In order to accurately identify all change points, they first fit the model where the tuning parameter's value is small, and subsequently apply screening criteria to remove any false positive change points. In their approach they did not derive any theoretical results with regards to change point identification consistency. Nor did they investigate the size or power of their proposed change point detection procedure. Furthermore, their method is heavily dependent on the choice of parameters. Our proposed procedure is free of tuning parameters and is theoretically rigorous.

While no methods in the existing literature are applicable to test (3.1) for high-dimensional functional data, it is also the case that the methods developed in Chapter 2 are not applicable for a few reasons. First, in Chapter 2 it was assumed that the number of repeated measurements is small. Numerical studies considered the finite sample performance when $T = 5$ and 8. A real data application was conducted where $T = 6$. Second, the asymptotic distribution of the test statistic and rate of convergence for the change point estimator were derived under an asymptotic setting in which p and n diverge but with T finite. For a large number of repeated measurements, as is the case with dense functional data, it will be more appropriate to consider an asymptotic setting in which p, n, and T diverge. Numerical simulation and real data applications should be based on theoretical results derived under this new asymptotic setting and not that considered in Chapter 2. Third, the computation complexity of the proposed procedure in Chapter 2 was not a concern for small values of n and T. The overall computation complexity of the change point detection procedure detailed in Chapter 2 is $O(pn^4T^6)$. To directly apply the procedure from Chapter 2 would be computationally impractical, if not impossible. Thus, in this chapter we aim to address these theoretical and computational challenges so our procedure is applicable to high-dimensional functional data.

In addition to testing the hypotheses of (3.1), we also develop a method to estimate unknown change points. In Chapter 2, the rate of convergence was established under an asymptotic setting where p and n diverge but with T finite. In this chapter we investigate the rate of convergence of the change point estimator where p, n, and T all diverge. Much of the research in change point identification considers the scenario with $n = 1$. For instance, Aue et al. (2009) considered a p-dimensional multivariate time series where T diverges but under the assumption that $p < T$. Wang et al. (2017) considered covariance matrix change point identification for T independent p-dimensional sub-Gaussian random vectors. They also require $p < T$. Dette et al. (2018) proposed a two-stage covariance change point identification procedure based on T independent sub-Gaussian random vectors. Their first step involves dimension reduction governed by a regularization parameter. In step two, they use a CUSUM-type statistic to estimate the locations of change points. Despite these recent advances, none of the aforementioned methods are applicable to identify change points among covariance matrices in high-dimensional functional data.

This chapter provides both theoretical and computational contributions to the field of statistics. From a theoretical perspective, a new asymptotic setting is considered, a setting suitable for high-dimensional functional data, in which n, p, and T diverge. For T diverging, the test statistic forms a stochastic process. The convergence of the finite-dimensional distributions is not sufficient for weak convergence of a stochastic process. Thus, we extend the finite-dimensional result to establish weak convergence of our proposed test statistic. Furthermore, the rate of convergence with respect to the change point estimator is now impacted by n, p, and T, as opposed to just n and p in Chapter 2. Our investigation reveals that the rate of convergence depends on the data dimension, sample size, number of repeated measurements, and signal-to-noise ratio. The change point identification estimator is shown to be consistent, provided the signal strength exceeds the noise. To our knowledge, the asymptotic framework in which n, p, and T all diverge has not previously been investigated with regards to change point identification among high-dimensional covariance matrices.

From a computation perspective, we improve the efficiency of methods developed in Chapter 2. This chapter considers T to be dense, so much of our attention is focused towards computation efficiency for those statistics that have high orders of T. We introduce two recursive relationships and computation efficient formulae to reduce the computation complexity from $O(pn^4T^6)$ to $O(pn^2T^4)$. A quantile approximation technique is shown to further decrease the complexity to the order of pn^2T^3. The approximation accuracy is demonstrated through simulation. These improvements are included in an R package, tecoma, which also affords an option for parallel computing. In the absence of these modifications, it would be impossible to apply our methods to fMRI data, or any high-dimensional data set with a large number of repeated measurements.

The remaining sections of this chapter are organized as follows. Section 3.2 details the statistical model and our basic setting. Section 3.3 introduces the measure from Chapter 2 along with the unbiased estimator that is a linear combination of U-type statistics. The test statistic's asymptotic distribution is derived under the asymptotic framework in which n, p, and T diverge. Computation consideration with regards to the statistics is provided in Section 3.4. Section 3.5 introduces an estimator to identify the locations of change points should we reject H_0 of (3.1). The estimator's rate of convergence is studied, and two procedures are detailed to estimate the locations of multiple change points. Sections 3.6 and 3.7 demonstrate the finite sample performance via simulation and investigate the brain's functional connectivity through a task-based fMRI data set, respectively. All proofs and technical details are provided in Section 3.8.

3.2 Model

Suppose we have n independent individuals that have p variables recorded at each of T identical time points. Let $Y_{it} = (Y_{it1}, \dots, Y_{itp})^{\mathrm{T}}$ be an observed p-dimensional random vector, where Y_{it} $(i = 1, \dots, n;\ t = 1, \dots, T)$ is independently and identically distributed for all n individuals. Assume Y_{it} follows a general factor model, where

$$Y_{it} = \mu_t + \Gamma_t Z_i, \tag{3.2}$$

and μ_t is a p-dimensional unknown mean vector, Γ_t is an unknown $p \times m$ matrix such that $m \geq pT$, and Z_i's are independent m-dimensional multivariate standard normal random vectors. Since $\mathrm{var}(Z_i) = I_m$, it follows that for the ith individual, $\mathrm{cov}(\Gamma_s Z_i, \Gamma_t Z_i) = \Gamma_s \Gamma_t^{\mathrm{T}}$. We define $\Gamma_s \Gamma_t^{\mathrm{T}}$ as C_{st} for different time points, s and t, and define $\Gamma_t \Gamma_t^{\mathrm{T}}$ as Σ_t. Thus, for the ith individual,

$$\mathrm{cov}(Y_{is}, Y_{it}) = \begin{cases} C_{st}, & s \neq t, \\ \Sigma_t, & s = t, \end{cases}$$

for all $s, t \in \{1, \ldots, T\}$. For individuals $i \neq j$, $\mathrm{cov}(Y_{is}, Y_{jt}) = 0$. By definition, C_{st} and Σ_t are $p \times p$ matrices for all $s, t \in \{1, \ldots, T\}$. No specific structure is required on covariance matrices C_{st} and Σ_t. Their generality allows us to capture the spatiotemporal dependence in and among the random vectors Y_{it} $(i = 1, \ldots, n;\ t = 1, \ldots, T)$. In the context of fMRI data, spatial dependence is present among neighboring voxels or nodes and is captured in both C_{st} and Σ_t. Temporal dependence exists for the same voxel or node across time points and is captured in matrix C_{st}.

3.3 Change point detection

We consider the measure, D_t $(t = 1, \ldots, T-1)$, defined in Chapter 2, where

$$D_t = \frac{1}{t(T-t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2})^2\}. \tag{3.3}$$

To simplify notation, let $t(T-t)$ be defined as $w(t)$. The choice of D_t is motivated by the fact that we can distinguish between H_0 and H_1 based on the maximum value of D_t for all $t \in \{1, \ldots, T-1\}$. Let $\mathcal{T} = \{1, \ldots, T-1\}$. Under H_0 of (3.1), $\max_{t \in \mathcal{T}} D_t = 0$, and under H_1, $\max_{t \in \mathcal{T}} D_t > 0$.

Our test statistic is constructed in the same manner as detailed in Chapter 2. We use a linear combination of U-type statistic estimators to create an unbiased estimator of D_t. Quantity D_t can be expressed as $D_t = w^{-1}(t) \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \{\mathrm{tr}(\Sigma_{s_1}^2) + \mathrm{tr}(\Sigma_{s_2}^2) - \mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2}) - \mathrm{tr}(\Sigma_{s_2}\Sigma_{s_1})\}$. An unbiased estimator for $\mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})$ is given by $U_{s_1 s_2}$, where

$$U_{s_1 s_2} = U_{s_1 s_2,0} - U_{s_1 s_2,1} - U_{s_2 s_1,1} + U_{s_1 s_2,2}, \tag{3.4}$$

and

$$
\begin{aligned}
P_n^2 U_{s_1 s_2,0} &= \sum_{i,j}^{\sim} (Y_{is_1}^{\mathrm{T}} Y_{js_2})^2, \\
P_n^3 U_{s_1 s_2,1} &= \sum_{i,j,k}^{\sim} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1}, \\
P_n^3 U_{s_2 s_1,1} &= \sum_{i,j,k}^{\sim} Y_{is_2}^{\mathrm{T}} Y_{js_1} Y_{js_1}^{\mathrm{T}} Y_{ks_2}, \\
P_n^4 U_{s_1 s_2,2} &= \sum_{i,j,k,l}^{\sim} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{ks_1}^{\mathrm{T}} Y_{ls_2}.
\end{aligned}
$$

In the above expressions, quantity $P_n^k = n!/(n-k)!$, and the $\sim$ summation notation represents the summation over mutually different indices. Thus, $\sum_{i,j,k}^{\sim}$ is defined as the summation over i, j, and k, such that $i \neq j$, $j \neq k$, and $k \neq i$. Therefore, an unbiased estimator of D_t is

$$
\begin{aligned}
\hat{D}_{nt} &= \frac{1}{w(t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} (U_{s_1 s_1} + U_{s_2 s_2} - U_{s_1 s_2} - U_{s_2 s_1}) \\
&= \frac{1}{w(t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{a,b=1}^{2} (-1)^{|a-b|} U_{s_a s_b}. \qquad (3.5)
\end{aligned}
$$

In this chapter we consider a different asymptotic framework than that of Chapter 2. Chapter 2 considered $p(n) \to \infty$ as $n \to \infty$, where p is a function of n. We now consider $p(n) \to \infty$ and $T(n) \to \infty$ as $n \to \infty$, where p and T are both functions of n. No specific functional form is required, and we do not require any specific relationships between p, T, and n. Thus, we allow for $p > n$ and $p > T$. To establish the limiting distribution of $\hat{D}_{nt}$, we assume Conditions 1 – 2 introduced in Section 2.3 along with the following two conditions. The notation $\sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}}$ is defined as $\sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{h_1=1}^{t} \sum_{h_2=t+1}^{T}$, and quantity V_{0t} is given by (3.6).

Condition 4. $\sum^{*}_{\substack{s_1,s_2,\\ h_1,h_2}} \mathrm{tr}^4(C_{s_u h_k} C_{s_v h_l}^{\mathrm{T}}) = o(V_{0t}^2)$*, for any* $u, k, v, l \in \{1, 2\}$.

Condition 5. *There exists a function* $\psi(k)$ *such that* $\psi(k) > 0$ *and* $\sum_{k=1}^{\infty} \psi(k) < \infty$. *For any* $s_1, s_2 \in \{1, \dots, T\}$, $\mathrm{tr}^2(C_{s_1 s_2} C_{s_2 s_1}) \asymp \psi(|s_1 - s_2|) \mathrm{tr}^2(\Sigma_{s_1} \Sigma_{s_2})$.

In Condition 5, $\asymp$ means of the same order. Thus, $f(s) \asymp g(s)$ implies there exists a constant c_1 such that $|f(s)| \leq c_1|g(s)|$, and there exists a constant c_2 such that $|g(s)| \leq c_2|f(s)|$ for all s in the real numbers. Condition 5 imposes mild requirements on the spatiotemporal structure. Condition 4 is also a mild condition. If no temporal dependence exists, then $V_{0t} = \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\sum_{u,v\in\{1,2\}} \mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v})$. Similarly, the left-hand side of Condition 4 is $\sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\sum_{u,v\in\{1,2\}} \mathrm{tr}^4(\Sigma_{s_u}\Sigma_{s_v})$. Furthermore, if all eigenvalues of Σ_t are bounded for all $t \in \{1, \ldots, T\}$, then $V_{0t}^2 \asymp \{t(T-t)p^2\}^2$. In comparison, the left-hand side of Condition 4 is of the order $t(T-t)p^4$. As a result, Condition 4 holds.

In Chapter 2, we derived the leading order variance of $\hat{D}_{nt}$, that is $\mathrm{var}(\hat{D}_{nt}) = \sigma_{nt}^2\{1 + o(1)\}$, where $\sigma_{nt}^2 = w^{-2}(t)(4V_{0t}/n^2 + 8V_{1t}/n)$, and

$$V_{0t} = \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{\substack{u,v,\\ k,l\in\{1,2\}}} (-1)^{|u-v|+|k-l|}\mathrm{tr}^2(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}), \tag{3.6}$$

$$V_{1t} = \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} \sum_{u,k\in\{1,2\}} (-1)^{|u-k|}\mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2})C_{s_uh_k}(\Sigma_{h_1} - \Sigma_{h_2})C_{s_uh_k}^{\mathrm{T}}\}. \tag{3.7}$$

The below theorem establishes the asymptotic distribution of $\hat{D}_{nt}$ under the asymptotic setting considered in this chapter.

Theorem 7. *Under Conditions 1 – 2, and 4, as $n \to \infty$,*

$$\sigma_{nt}^{-1}\left(\hat{D}_{nt} - D_t\right) \xrightarrow{d} N(0,1),$$

where $\sigma_{nt}^2 = w^{-2}(t)(4V_{0t}/n^2 + 8V_{1t}/n)$ and V_{0t} and V_{1t} are given in (3.6) and (3.7), respectively.

Under the null hypothesis, it follows that $\sigma_{nt,0}^{-1}\hat{D}_{nt} \to N(0,1)$ in distribution, where $\sigma_{nt,0}^2 = w^{-2}(t)(4V_{0t}/n^2)$ and only Conditions 1 and 4 are required. To formulate an appropriate test procedure free of tuning parameters, consider the test statistic, $\mathcal{M}_n$, of Chapter 2, where

$$\mathcal{M}_n = \max_{t\in\mathcal{T}} \hat{\sigma}_{nt,0}^{-1}\hat{D}_{nt}, \tag{3.8}$$

and $\hat{\sigma}_{nt,0}$ is a plug-in estimator for $\sigma_{nt,0}$. Methods to construct $\hat{\sigma}_{nt,0}$ were detailed in Chapter 2. The following theorem establishes the asymptotic distribution of $\mathcal{M}_n$ under the setting where n, p, and T all diverge.

Theorem 8. *Under Conditions 1, 4, and 5, H_0 of (3.1), and as $n \to \infty$, $\mathcal{M}_n \xrightarrow{d} \max_{t \in \mathcal{T}} Z_t$, where Z_t is a Gaussian process with mean 0 and covariance R_z.*

We assume that as $n \to \infty$, $R_{n,z}$ converges to R_z, where $R_{n,z}$ is a correlation matrix with (t,q) component defined as $R_{n,tq} = \operatorname{corr}(\hat{D}_{nt}, \hat{D}_{nq})$. The leading order of the $\operatorname{cov}(\hat{D}_{nt}, \hat{D}_{nq})$ is $w^{-1}(t)w^{-1}(q)(4V_{0,tq}/n^2)$, where

$$V_{0,tq} = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{h_1=1}^{q} \sum_{h_2=q+1}^{T} \sum_{\substack{u,v,\\ k,l \in \{1,2\}}} (-1)^{|u-v|+|k-l|} \operatorname{tr}^2(C_{s_u h_k} C_{s_v h_l}^{\mathrm{T}}). \tag{3.9}$$

In order to perform an α-level hypothesis test for (3.1), we must approximate $R_{n,z}$ and thus require an estimator for $V_{0,tq}$. In Chapter 2, an unbiased estimator for $\operatorname{tr}(C_{s_u h_k} C_{s_v h_l}^{\mathrm{T}})$ was given as a linear combination of U-type statistics. Let $\hat{R}_{n,tq}$ be an estimator for the (t,q) component of $R_{n,z}$.

Let $W = \max_{t \in \mathcal{T}} Z_t$, where Z_t is a Gaussian process with mean 0 and covariance R_z, and define W_α as the quantity such that $\operatorname{pr}(W > W_\alpha) = \alpha$. By Theorem 8, $\mathcal{M}_n \to W$ in distribution, and an α-level test rejects the null hypothesis in (3.1) if $\mathcal{M}_n > W_\alpha$. However, there is no simple and computation efficient approach to obtain W_α. The random variable W depends on R_z. Chapter 2 proposed a procedure to approximate quantile W_α on the basis of computing $\hat{R}_{n,tq}$ for each $t, q \in \{1, \ldots, T-1\}$. However, the computation complexity of this approach, in terms of T, is at the order of T^4 for each component. Therefore, total complexity is at the order of T^6 to compute $\hat{R}_{n,z}$. As a result, it is not feasible to compute all components when T is large.

As an attempt to alleviate this burden, we can further approximate the distribution of W by a Gumbel distribution. Under additional assumptions and if T diverges, then

$$\operatorname{pr}(\mathcal{M}_n \le \sqrt{2\log(T) - \log\{\log(T)\} + x}) \to \exp\{-(2\sqrt{\pi})^{-1} \exp(-x/2)\}.$$

Accordingly, an α-level quantile is defined as $\sqrt{2\log(T) - \log\{\log(T)\} + x_\alpha}$, where $x_\alpha = -2\log\{-2\sqrt{\pi}\log(1-\alpha)\}$. However, the rate of convergence is at the order of $\log(T)$, which is slow. In addition, our simulation experiments demonstrated that the size of the test was not well controlled at the nominal level. Moreover, using an extreme value-type distribution does not eliminate the need to compute $\hat{\sigma}_{nt,0}$ for all $t \in \mathcal{T}$. That overall cost in terms of T is at the order T^5. Hence, we carefully consider an approximation procedure in Section 3.4 that improves efficiency and maintains accuracy.

3.4 Computation of the proposed statistics

The computation complexity for the change point detection procedure is at the order of pn^4T^6. To reduce the complexity, we re-formulate some of the statistics introduced in Section 3.3 in a computation optimal manner.

The computation complexity of $U_{s_1s_2}$ is at the order of n^4 due to term $U_{s_1s_2,2}$. In addition, term $U_{s_1s_2,1}$ has computation complexity at the order of n^3. To save computation cost, we can rewrite $U_{s_1s_2,1}$ and $U_{s_1s_2,2}$ defined in (2.4) in a computation efficient form as follows. First, we consider $U_{s_1s_2,1}$, which can be rewritten as

$$P_n^3 U_{s_1s_2,1} = \sum_{j=1}^{n}\Big(\sum_{i=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2}\Big)^2 - \sum_{i,j=1}^{n}(Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - 2\sum_{k\neq j=1}^{n} Y_{js_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1}. \tag{3.10}$$

Therefore, the computation complexity of $U_{s_1s_2,1}$ regarding the sample subjects is at the order of n^2, not n^3. To write $U_{s_1s_2,2}$ in a computation efficient form, we first define $V_{s_1s_2,1} = (1/P_n^3)\,\tilde{\sum}_{i,j,k}\, Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{ks_2}$. Similar to $U_{s_1s_2,1}$, we can write $V_{s_1s_2,1}$ as

$$\begin{aligned} P_n^3 V_{s_1s_2,1} = {} & \sum_{j=1}^{n}\Big(\sum_{i=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2}\Big)\Big(\sum_{i=1}^{n} Y_{is_2}^{\mathrm{T}} Y_{js_1}\Big) - \sum_{i,j=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{is_2} \\ & - \sum_{k\neq j=1}^{n} Y_{js_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{ks_2} - \sum_{i\neq j=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{js_2}. \end{aligned}$$

The computation complexity of $V_{s_1s_2,1}$ regarding the sample subjects is also at the order of

n^2. Finally, we can write $U_{s_1 s_2,2}$ as

$$P_n^4 U_{s_1 s_2,2} = \Big(\sum_{i \neq j=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} \Big)^2 - P_n^3 (U_{s_1 s_2,1} + U_{s_2 s_1,1} + 2V_{s_1 s_2,1}) - P_n^2 U_{s_1 s_2,0} \\ - \sum_{i \neq j=1}^{n} (Y_{is_1}^{\mathrm{T}} Y_{js_2})(Y_{is_2}^{\mathrm{T}} Y_{js_1}). \tag{3.11}$$

Based on the above expression for $P_n^4 U_{s_1 s_2,2}$, we can also see that the computation complexity of $U_{s_1 s_2,2}$ regarding the sample subjects is at the order of n^2. In summary, the computation cost of the proposed statistic $U_{s_1 s_2}$ with regard to sample subjects is at the order of n^2. These computation efficient expressions can be derived in a similar manner for $U_{s_u s_v, h_k h_l}$, the term used as a plug-in estimator to primarily compute $\hat{R}_{n,tq}$.

The computation complexity of $\hat{D}_{nt}$ in (3.5) in terms of T is at the order T^3. To reduce the complexity in terms of T, we write $\hat{D}_{nt}$ recursively. Let $f(s_1, s_2) = (U_{s_1 s_1} + U_{s_2 s_2} - U_{s_1 s_2} - U_{s_2 s_1})$ such that $s_1, s_2 \in \{1, \ldots, T\}$. By definition, it follows that for $t \geq 2$,

$$\hat{D}_{nt} = \frac{w(t-1)}{w(t)} \hat{D}_{n(t-1)} - w^{-1}(t) \sum_{k=1}^{t-1} f(k,t) + w^{-1}(t) \sum_{k=t+1}^{T} f(t,k). \tag{3.12}$$

When $t = 1$, the computation complexity of $\hat{D}_{n1}$ is at the order of T. Therefore by (3.12), for each $t \in \{1, \ldots, T-1\}$ the computation complexity in terms of T is at the order T. Since we compute $\hat{D}_{nt}$ for all $t \in \{1, \ldots, T-1\}$, the total computation complexity in terms of T is at the order of T^2 rather than T^3. As a result, the overall computation complexity to compute $\hat{D}_{nt}$ for all $t \in \{1, \ldots, T-1\}$ is at the order of pn^2T^2. Parallel computing can further decrease the computation time.

The greatest cost in terms of computation is due to $\hat{R}_{n,tq}$ for all $t, q \in \mathcal{T}$, where the complexity is at the order of pn^2T^6 provided (3.10) and (3.11) are applied. To reduce the complexity, we express $\hat{R}_{n,tq}$ recursively. Let

$$g(s_1, h_1, s_2, h_2) = \sum_{\substack{u,v,\\ k,l \in \{1,2\}}} (-1)^{|u-v|+|k-l|} U^2_{s_u s_v h_k h_l},$$

$$h(t,q) = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{h_1=1}^{q} \sum_{h_2=q+1}^{T} g(s_1, h_1, s_2, h_2).$$

Thus, $n^2 w(t) w(q) \hat{V}_{0,tq}/4 = h(t,q)$. Suppose the quantity $h(t, q-1)$ is known for $t \in \{1, \ldots, T-2\}$ and $q \in \{2, \ldots, 2-1\}$. For a fixed t,

$$h(t,q) = h(t,q-1) - \sum_{j=t}^{q-1} \sum_{k=t+1}^{T} g(t,j,k,q) + \sum_{j=t+1}^{T} \sum_{k=q+1}^{T} g(t,q,j,k). \tag{3.13}$$

An analogous recursive formula can be derived to traverse a fixed column where $h(t-1,q)$ is known and we want to compute $h(t,q)$. Based on the recursive formula in (3.13), the computation complexity in terms of T is at T^2. By the definition of $\hat{R}_{n,tq}$, $\hat{R}_{n,1,1}$, $\hat{R}_{n,1,T-1}$, $\hat{R}_{n,T-1,T-1}$ can each be computed at the computation complexity in terms of T at the order T^2. Therefore, based on (3.13) and the fact that we must compute $\hat{R}_{n,tq}$ for all $t < q$, the overall computation complexity for $\hat{R}_{n,z}$ is at the order np^2T^4. Despite this reduction, the complexity can further be improved via linear interpolation on a sparse form of $\hat{R}_{n,z}$.

Rather than compute $\hat{R}_{n,tq}$ for all $t, q \in \{1, \ldots, T-1\}$, we can compute $h = (b + I)$ off-diagonals of the matrix and interpolate the remaining values. Let b be the number of consecutive off-diagonals immediately following the main diagonal, and let I be the number of off-diagonals computed at a fixed interval after the b consecutive off-diagonals. Let $\text{diag}(\hat{R}_{n,1,d+1})$ be the dth off-diagonal, where $d \in \{1, \ldots, T-2\}$. For an efficient approximation of $\hat{R}_{n,z}$, first compute $\hat{R}_{n,1,1}$. Next, apply (3.13) to compute $\text{diag}(\hat{R}_{n,1,2}), \ldots, \text{diag}(\hat{R}_{n,1,b})$ for the corresponding b off-diagonals. Lastly, apply formula (3.13) to compute $\text{diag}(\hat{R}_{n,1,I_1}), \ldots, \text{diag}(\hat{R}_{n,1,I_I})$ that correspond to the I off-diagonals at a fixed interval. Each of these I off-diagonals has an initial computation in terms of T at the order T^3. Parallel processing can be utilized to start each off-diagonal's computation independently. The overall complexity in terms of T, will be at order hT^3 to obtain a sparse version of

$\hat{R}_{n,z}$. Linear interpolation is then used to estimate the components not computed. Based on our simulations, linear interpolation results in a negligible loss in power, and the size remains near the nominal level. Full simulation results for the linear interpolation method are available in Section 3.6.

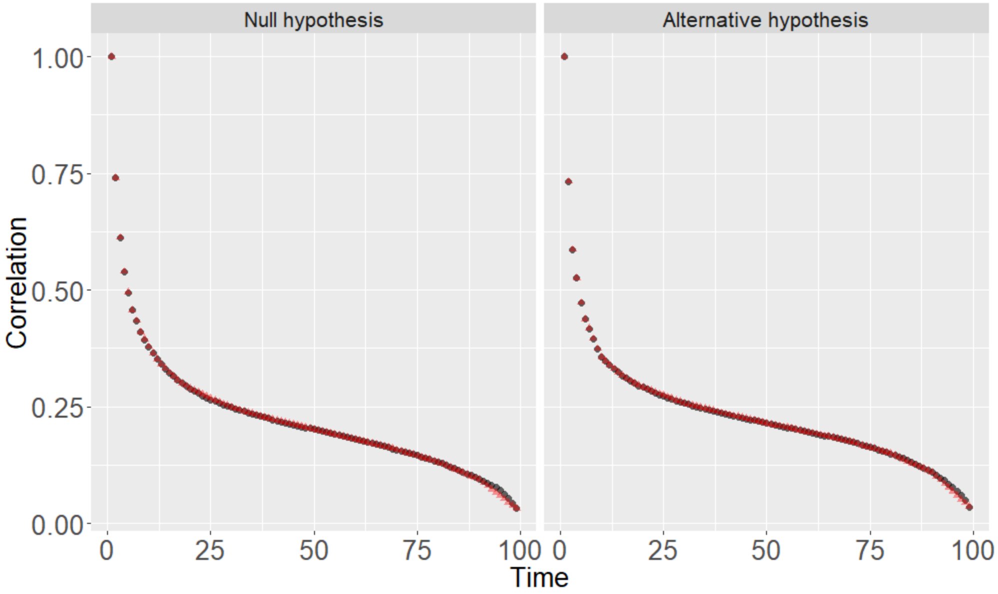

Figure 3.1: Accuracy of linear interpolation for $\hat{R}_{n,tq}$. Black circles represent $\hat{R}_{n,1q}$ for all $q \in \{1, \dots, T-1\}$. Red triangles represent the corresponding interpolated values.

Figure 3.1 illustrates $\hat{R}_{n,1,q}$ for all $q \in \{1, \dots, T-1\}$ and the corresponding interpolated values based on parameters $b = 20$ and $I = 8$. The fixed interval for off-diagonals was set to ten. The accuracy of the linear interpolation is evident under the null and alternative hypotheses. Therefore, the computation complexity in terms of T for $\hat{R}_{n,tq}$ can be reduced from T^4 to hT^3. In Chapter 2, the overall computation complexity to approximate the quantile was at the order pn^4T^6. From the recursive formulae and estimation procedure via linear interpolation, the overall computation complexity to estimate $R_{n,z}$ is reduced to pn^2T^3, and thus, making our change point detection procedure applicable to high-dimensional functional data.

3.5 Change point identification

If the data leads us to the conclusion to reject H_0 of (3.1), then a second task is to identify the time points where changes exist among the T high-dimensional covariance matrices. First, consider the case with only one change point. Let τ be the time point where the single change point exists. Define $\hat{\tau}$ as an estimator for the change point's location, where

$$\hat{\tau} = \arg\max_{t \in \mathcal{T}} \hat{D}_{nt}, \tag{3.14}$$

and $\mathcal{T} = \{1, \ldots, T-1\}$. The form of the estimator is motivated by Theorem 4, which states that D_t is maximized at the time point $t = \tau$ when a single change point exists at τ. Consider the hypotheses

$$\begin{aligned} H_0 :& \quad \Sigma_1 = \cdots = \Sigma_T \qquad \text{versus} \\ H_1^* :& \quad \Sigma_1 = \cdots = \Sigma_\tau \neq \Sigma_{\tau+1} = \cdots = \Sigma_T. \end{aligned} \tag{3.15}$$

The following theorem establishes the rate of convergence for $\hat{\tau}$.

Theorem 9. *Assume that H_1^* of (3.15) is true. Also, assume that as $T \to \infty$, $\tau/T \to \omega$, a constant. Under Conditions 1 – 2, and 4, it follows that as $n \to \infty$,*

$$\hat{\tau} - \tau = O_p\Big\{\frac{\nu_{\max}\sqrt{\log(T)}}{n\Delta_p}\Big\}, \tag{3.16}$$

where $\nu_{\max} = \max_{t \in \mathcal{T}} \max\left(\sqrt{V_{0t}}, \sqrt{nV_{1t}}\right)$ and $\Delta_p = \mathrm{tr}\{(\Sigma_1 - \Sigma_T)^2\}$.

Theorem 9 demonstrates that the change point estimator, $\hat{\tau}$, is consistent for high-dimensional functional data, provided that $\Delta_p/\nu_{\max} \gg \sqrt{\log(T)}/n$. Quantity Δ_p can be interpreted as the signal, and quantity $\nu_{\max}$ can be interpreted as the noise. Thus, if $\nu_{\max}\sqrt{\log(T)}/(n\Delta_p) \to 0$, $\hat{\tau}$ is a consistent estimator for τ.

To investigate the impact of n, p, and T on the rate of convergence of $\hat{\tau}$ we consider each in turn. Assume p and T are fixed as $n \to \infty$. As a result, the rate of convergence for $\hat{\tau} - \tau$ is $O_p(1/\sqrt{n})$ since Δ_p, $\sqrt{\log(T)}$, $\sqrt{V_{0t}}$, and $\sqrt{V_{1t}}$ are held constant. If we assume T is fixed as n and p diverge, then the rate of convergence is the same as that proved in Theorem 4. This rate can be faster than $1/\sqrt{n}$ depending on the contributions of Δ_p and $\nu_{\max}$. Next, if we assume p is fixed as n and T diverge, then $\hat{\tau} - \tau = O_p\Big\{\nu_{\max}\sqrt{\log(T)}/\sqrt{n}\Big\}$. Depending on the relationship between Δ_p and $\nu_{\max}$, the rate of convergence can be much faster than $\sqrt{\log(T)}/\sqrt{n}$ as p, T, and n all diverge. As p increases, $\Sigma_1 - \Sigma_T$ can possibly contain more nonzero components, so Δ_p could get larger. However, as p and T increase, $\nu_{\max}$ increases. Therefore, if $\nu_{\max}$ does not dominate Δ_p, we obtain a faster rate of convergence than $\sqrt{\log(T)}/\sqrt{n}$. Despite the fact that the estimator in (3.14) is the same as that proposed in Chapter 2, the rate of convergence for the estimator is very different with regards to the asymptotic framework in which $n \to \infty$, $p \to \infty$, $T \to \infty$.

Assume H_1 of (3.1) is true for multiple change points. First, we introduce two procedures to identify the locations of multiple change points, and then introduce a theorem with regards to the consistency of estimating multiple change points.

Let $\mathcal{Q} = \{1 \leq \tau_1 < \cdots < \tau_q < T\}$ be the collection of all the true q change points, and let $\hat{\mathcal{Q}}$ be the estimated set of change points. We make use of the notation in Chapter 2 and define for time points $t_1 < t_2$, $S[t_1, t_2]$ is the statistic S calculated based on the data in time interval t_1 through t_2. For example, $\nu_{\max}[t_1, t_2]$ is the quantity based on data between t_1 and t_2. To identify multiple change points we apply binary segmentation (Venkatraman 1992). The binary segmentation algorithm is detailed follows.

Step 1: Compute $\mathcal{M}_n$ and compare it with W_α. If $\mathcal{M}_n > W_\alpha$, then $\hat{\kappa} = \arg\max_{t \in \mathcal{T}} \hat{D}_{nt}$ is the estimated change point, and set $\hat{\kappa} = \hat{\tau}_1$ so $\hat{\mathcal{Q}} = \{\hat{\tau}_1\}$. Partition the full data set into two intervals: $[1, \hat{\kappa}]$ and $[\hat{\kappa}+1, T]$ and proceed to step 2. However, if $\mathcal{M}_n \leq W_\alpha$, then no change points exist.

Step 2: Perform the detection procedure to test (3.1) using $Y[1, \hat{\kappa}]$ and $Y[\hat{\kappa}+1, T]$. If H_0 is rejected based on $Y[1, \hat{\kappa}]$, then identify $\hat{\kappa}_1 = \arg\max_{t \in [1, \hat{\kappa}]} \hat{D}_{nt}[1, \hat{\kappa}]$ as a change point. Since $\hat{\kappa}_1 < \hat{\tau}_1$, set $\hat{\tau}_1 = \hat{\kappa}_1$ and $\hat{\tau}_2 = \hat{\kappa}$ so $\hat{\mathcal{Q}} = \{\hat{\tau}_1, \hat{\tau}_2\}$. Partition the data $Y[1, \hat{\kappa}]$ into two intervals: $[1, \hat{\kappa}_1]$ and $[\hat{\kappa}_1 + 1, \hat{\kappa}]$. If H_0 is not rejected, then no change points exist in the interval $[1, \hat{\kappa}]$. Repeat this procedure for the data based on interval $[\hat{\kappa} + 1, T]$. Set $\hat{\mathcal{Q}}$ is then updated to contain the ordered change points. If no change points are detected in either interval, then stop, as $\hat{\kappa}$ is the only change point that exists.

Step 3: If a change point is identified in at least one interval in step 2, repeat step 2 until no further change points are detected. At each step update and order set $\hat{\mathcal{Q}}$.

At the conclusion of the binary segmentation procedure we can partition the interval $[1, T]$ so each sub-interval will consist of end points from the set $\{1, \hat{\mathcal{Q}}, T\}$. For example, if no change point is identified, then the single interval is $[1, T]$. If a single change point is identified at $\hat{\tau}$, then two intervals where no change points exist are $[1, \hat{\tau}]$ and $[\hat{\tau}, T]$.

The computation time to identify multiple change points exceeds the time to detect the existence of change points. If parallel computing is available, then the computation time required to identify multiple change points can be improved via a more efficient identification procedure when compared to the steps outlined for binary segmentation. The improvement stems from the fact that the time to compute $\hat{D}_{nt}$ is less than the time required to test (3.1) for a given time interval. An efficient parallel procedure is detailed below.

Step 1: Perform binary segmentation by partitioning at $\arg\max_{t \in I_t} \hat{D}_{nt}[I_t]$, where I_t is the considered time interval of data. Change point detection is not performed at this step. Binary segmentation continues until all intervals are either of the form $[a, a]$ or $[a, a+1]$, where $a \in \mathcal{T}$. Suppose there exist N total intervals at the conclusion of binary segmentation.

Step 2: For all N intervals of length at least one, apply the change point detection procedure to test (3.1) in parallel. If H_0 is rejected for a given interval, then a change point exists and is estimated at the point $\arg\max_{t \in I_t} \hat{D}_{nt}[I_t]$ from step 1. Update $\hat{\mathcal{Q}}$ for each identified change point.

Hence, the computation time required to identify multiple change points will only slightly exceed the time to perform change point detection on the longest interval $[1, T]$.

To establish the consistency of $\hat{\mathcal{Q}}$ we first define some notation. Let I_t be a time interval such that $I_t = [\tau_a + 1, \tau_b]$, where $a+1 < b$ such that $a \in \{0, \ldots, q-1\}$ and $b \in \{2, \ldots, q+1\}$. Define $\tau_0 = 0$ and $\tau_{q+t} = T$. Thus, I_t is an interval with at least one change point. Assume the smallest maximum signal-to-noise ratio among all segments I_t is as defined in Chapter 2, where $\min_{I_t} \max_{\tau_s \in I_t} \sigma^{-1}_{n\tau_s,0}[I_t] D_{\tau_s}[I_t]$ is denoted as mSNR.

Theorem 10. *Assume that* τ_k/T *converges to* ω_k *as* T *diverges,* $W_{\alpha_n} = o(\mathrm{mSNR})$*, and for any interval* I_t*,* $\nu_{\max}[I_t]\sqrt{\log(T)}/(n\Delta_p[I_t]) \to 0$*. Furthermore, assume* $\alpha_n \to 0$*. Therefore, under Conditions 1 – 2, and 4, as* $n \to \infty$*,* $\hat{\mathcal{Q}} \to \mathcal{Q}$ *in probability.*

In the existence of change points, the assumption that $W_{\alpha_n} = o(\mathrm{mSNR})$ ensures the consistency of the proposed test at each phase of binary segmentation. In the absence of change points, the assumption that $\alpha_n \to 0$ ensures no change points will be detected and binary segmentation will stop on the given interval. The assumption that $\nu_{\max}[I_t]\sqrt{\log(T)}/(n\Delta_p[I_t]) \to 0$ ensures that in the existence of change points, the estimator is consistent.

3.6 Simulation studies

In this section, we present multiple simulation studies to demonstrate the performance of the change point detection and identification procedures in a large p, large T, and small n setting. All data were generated from a multivariate linear process,

$$Y_{it} = \sum_{h=0}^{L} A_{t,h}\xi_{i(t-h)} \quad (i = 1, \ldots, n;\ t = 1, \ldots, T), \tag{3.17}$$

where $A_{t,h}$ is a $p \times p$ matrix, and $\xi_{i(t-h)}$ are p-dimensional multivariate normally distributed random vectors with mean 0 and covariance I_p. The data generation scheme given by (3.17) permits spatial and temporal dependence. Let $t \geq s$. By definition of Y_{it} in (3.17),

$$\operatorname{cov}(Y_{it}, Y_{is}) = \begin{cases} \displaystyle\sum_{h=t-s}^{L} A_{t,h} A_{s,h-(t-s)}^{\mathrm{T}}, & t-s \leq L; \\ 0, & t-s > L. \end{cases}$$

Spatial dependence occurs among the vector Y_{it} for a given time point t. Temporal dependence exists among $\{Y_{it}\}_{t=1}^{T}$ at different time points and is governed by the simulation parameter L.

In the simulation studies, we set $n = 40, 50$ and 60, and $p = 500, 750$ and 1000. The number of repeated measurements, T, was set to be 50 and 100. For change point identification we considered an additional case with $T = 150$. The simulation parameter $L = 3$.

Simulation results reported in Tables 3.1 – 3.4 were based on 500 simulation replications, and simulation results in Tables 3.5 and 3.6 were based on 100 simulation replications.

The spatial and temporal dependence incorporated in (3.17) depends on the choice of matrices $A_{t,h}$. First, we define matrices $A_{t,h}$ for the testing simulation to demonstrate the size and power of the proposed test procedure. Later, matrices $A_{t,h}$ will be defined for the change point identification simulation.

Let τ_1 be the true underlying change point among the covariance matrices such that $\tau_1 = \lfloor T/2 \rfloor$, where $\lfloor x \rfloor$ is the floor function. Define two matrices, B_1 and B_2, such that

$$B_1 = \Big\{(0.6)^{|i-j|} I(|i-j| < p/5)\Big\},$$
$$B_2 = \Big\{(0.6+\delta)^{|i-j|} I(|i-j| < p/5)\Big\},$$

where (i, j) represents the ith row and jth column of the $p \times p$ matrices B_1 and B_2. Thus, for $h \in \{0, \ldots, 3\}$

$$A_{t,h} = \begin{cases} B_1, & t \in \{1, \ldots, \tau_1\}; \\ B_2, & t \in \{\tau_1 + 1 \ldots, T\}. \end{cases}$$

Parameter δ in B_2 governs the signal strength in terms of how different the covariance matrices are before and after the change point at time τ_1. When $\delta = 0$, $B_1 = B_2$ and $A_{t,h}$

is the same for all t, and the null hypothesis is true. If $\delta > 0$, then the null hypothesis is false, and τ_1 is the true covariance change point. For the change point detection simulation, δ was set to $0.00, 0.025, 0.05$ and 0.10.

Table 3.1: Empirical size and power of the proposed test, percentages of simulation replications that reject the null hypothesis

		$T = 50$			$T = 100$		
			p			p	
δ	n	500	750	1000	500	750	1000
	40	4.4	4.6	3.8	3.6	5.4	4.4
0(size)	50	4.8	4.0	3.6	2.0	4.6	4.0
	60	3.8	4.2	2.8	5.4	3.6	5.6
	40	13.4	13.4	10.8	18.0	19.0	18.0
0.025	50	17.0	19.2	17.0	30.6	27.2	30.4
	60	26.4	26.0	27.4	47.0	41.6	41.6
	40	96.0	97.0	98.0	100	100	100
0.05	50	100	100	100	100	100	100
	60	100	100	100	100	100	100
	40	100	100	100	100	100	100
0.10	50	100	100	100	100	100	100
	60	100	100	100	100	100	100

Table 3.1 demonstrates the empirical size and power of the proposed test procedure. The size is well controlled at the nominal level of 0.05 for all values of n, p, and T. For a fixed p and T, as n increases the power increases. Likewise, as δ increases, the power of the change point detection procedure increases. For a fixed n and p, the power increases as T increases. These relationships are further elucidated when simulation results from Table 2.1 under Setting (I) in Section 2.5 are considered. For example, when $n = 40$, $p = 500$, and $\delta = 0.05$, we observe that the power of the test is 21.4, 35.6, 96.0, and 100 as T is 5, 8, 50, and 100, respectively.

Table 3.2: Empirical size and power of the proposed test for $T = 100$, percentages of simulation replications that reject the null hypothesis, quantile computed from a correlation matrix that used linear interpolation. The first 5 off-diagonals were computed exactly as well as the last w components for each row

			$w = 5$			$w = 10$			$w = 20$	
			p			p			p	
δ	n	500	750	1000	500	750	1000	500	750	1000
0(size)	40	3.4	4.8	4.2	3.4	4.8	4.2	3.4	5.2	4.2
	50	2.0	4.6	3.8	2.0	4.6	4.0	2.0	4.6	4.0
	60	4.8	3.2	5.0	4.8	3.2	5.0	5.2	3.8	5.6
0.025	40	17.8	19.0	17.6	17.8	19.0	17.6	17.8	19.0	17.6
	50	30.8	26.2	30.2	30.8	26.6	30.2	30.8	26.6	30.2
	60	46.6	40.8	41.0	46.6	40.8	41.0	46.6	41.2	41.0
0.05	40	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100
	60	100	100	100	100	100	100	100	100	100
0.10	40	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100
	60	100	100	100	100	100	100	100	100	100

Table 3.3: Empirical size and power of the proposed test for $T = 100$, percentages of simulation replications that reject the null hypothesis, quantile computed from a correlation matrix that used linear interpolation. The first 10 off-diagonals were computed exactly as well as the last w components for each row

			$w = 5$			$w = 10$			$w = 20$	
			p			p			p	
δ	n	500	750	1000	500	750	1000	500	750	1000
0(size)	40	3.4	5.0	4.2	3.4	5.0	4.2	3.4	5.2	4.2
	50	2.0	4.6	4.0	2.0	4.6	4.0	2.0	4.6	4.0
	60	4.8	3.2	5.0	4.8	3.4	5.0	5.2	3.8	5.6
0.025	40	18.0	19.0	17.6	18.0	19.0	17.6	18.0	19.0	17.6
	50	30.8	26.6	30.2	30.8	26.6	30.2	30.8	26.8	30.2
	60	46.6	40.8	41.0	46.6	40.8	41.0	46.6	41.2	41.0
0.05	40	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100
	60	100	100	100	100	100	100	100	100	100
0.10	40	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100
	60	100	100	100	100	100	100	100	100	100

Table 3.4: Empirical size and power of the proposed test for $T = 100$, percentages of simulation replications that reject the null hypothesis, quantile computed from a correlation matrix that used linear interpolation. The first 20 off-diagonals were computed exactly as well as the last w components for each row

		$w = 5$			$w = 10$			$w = 20$		
			p			p			p	
δ	n	500	750	1000	500	750	1000	500	750	1000
0(size)	40	3.6	5.2	4.2	3.6	5.2	4.4	3.6	5.2	4.4
	50	2.0	4.6	4.0	2.0	4.6	4.0	2.0	4.6	4.0
	60	5.2	3.4	5.6	5.2	3.4	5.6	5.2	3.8	5.6
0.025	40	18.0	19.0	17.6	18.0	19.0	17.8	18.0	19.0	18.0
	50	30.6	26.8	30.2	30.6	26.8	30.4	30.6	27.0	30.4
	60	46.8	40.8	41.0	47.0	41.4	41.4	46.6	41.4	41.6
0.05	40	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100
	60	100	100	100	100	100	100	100	100	100
0.10	40	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100
	60	100	100	100	100	100	100	100	100	100

Tables 3.2 – 3.4 demonstrate the empirical size and power of the proposed test procedure using a modification of the quantile approximation procedure introduced in Section 3.4. Rather than compute $\hat{R}_{n,tq}$ for all $t, q \in \{1, \ldots, T-1\}$, we compute the first b off-diagonals and the last w columns of $\hat{R}_{n,tq}$. The remaining values were imputed via linear interpolation. Figure 3.1 demonstrates the accuracy of this linear interpolation procedure. Simulations considered $b = 5$, 10 and 20, and $w = 5$, 10 and 20. Based on our simulation results, there is only a minimal loss in power when compared to computing all components of $\hat{R}_{n,tq}$. Furthermore, the size of the test is well maintained at the nominal level of 0.05.

To evaluate the performance of the change point identification procedure through binary segmentation, consider two change points: τ_1 and τ_2. Let $\tau_1 = \lfloor T/2 \rfloor$, and let $\tau_2 = \tau_1 + 2$.

Define three matrices, B_1, B_2, and B_3, such that

$$\begin{aligned} B_1 &= \Big\{(|i-j|+1)^{-2} I(|i-j| < p/5)\Big\}, \\ B_2 &= \Big\{(|i-j|+\delta+1)^{-2} I(|i-j| < p/5)\Big\}, \\ B_3 &= \Big\{(|i-j|+2\delta+1)^{-2} I(|i-j| < p/5)\Big\}, \end{aligned}$$

where (i, j) represents the ith row and jth column of the $p \times p$ matrices B_1, B_2, and B_3. Thus, for $h \in \{0, \ldots, 3\}$

$$A_{t,h} = \begin{cases} B_1, & t \in \{1, \ldots, \tau_1\}; \\ B_2, & t \in \{\tau_1 + 1 \ldots, \tau_2\}; \\ B_3, & t \in \{\tau_2 + 1 \ldots, T\}. \end{cases}$$

When $\delta = 0$, the null hypothesis is true, and $A_{t,h}$ is the same for all $t \in \{1, \ldots, T\}$. Since our purpose is to demonstrate the finite sample accuracy of change point identification, we do not consider a null hypothesis setting in which $\delta = 0$. The values of δ were selected to be 0.15, 0.25, and 0.35.

Two measures were considered to evaluate the change point identification procedure's efficacy: average true positives and average true negatives. For each simulation replication there exists two true change points at time points τ_1 and τ_2, and there exists $T - 3$ time points where no change point exists. The average true positives are defined as the average number of correctly-identified change points among 100 simulation replications. Similarly, the average true negatives are defined as the average number of correctly-identified time points where no covariance change exists among 100 simulation replications.

Table 3.5 provides the efficacy of the binary segmentation procedure in the large p, large T, and small n setting. For fixed p, n, and T, the average true positives and average true negatives approach two and $T - 3$, respectively, as δ increases. As the sample size increases, the average true positives and average true negatives approach their optimal values. Table 3.6 contains the corresponding standard errors for the measures in Table 3.5.

Table 3.5: Average true positives and average true negatives for identifying multiple change points using the proposed binary segmentation method. The maximum number of true positives for a given replication is 2. The maximum number of true negatives for a given replication is $T-3$

			δ=0.15		δ=0.25		δ=0.35	
T	p	n	ATP	ATN	ATP	ATN	ATP	ATN
50	500	40	1.20	46.76	1.68	46.48	1.97	46.62
		50	1.41	46.68	1.91	46.42	2.00	46.63
		60	1.57	46.58	1.98	46.52	2.00	46.61
	750	40	1.30	46.78	1.77	46.51	2.00	46.59
		50	1.33	46.66	1.95	46.53	2.00	46.70
		60	1.57	46.58	1.99	46.53	2.00	46.64
	1000	40	1.27	46.76	1.81	46.61	1.95	46.59
		50	1.48	46.67	1.95	46.58	2.00	46.76
		60	1.65	46.51	1.99	46.69	2.00	46.59
100	500	40	1.27	96.75	1.74	96.56	1.98	96.54
		50	1.31	96.67	1.92	96.54	2.00	96.44
		60	1.62	96.70	1.99	96.56	2.00	96.46
	750	40	1.22	96.76	1.85	96.59	1.98	96.54
		50	1.33	96.59	1.96	96.51	2.00	96.54
		60	1.60	96.59	1.99	96.55	2.00	96.42
	1000	40	1.20	96.80	1.74	96.52	1.98	96.59
		50	1.34	96.64	1.90	96.50	2.00	96.49
		60	1.59	96.50	2.00	96.58	2.00	96.44
150	500	40	1.19	146.76	1.73	146.53	1.97	146.48
		50	1.34	146.68	1.95	146.55	2.00	146.40
		60	1.54	146.51	2.00	146.57	2.00	146.53
	750	40	1.16	146.84	1.73	146.58	1.97	146.46
		50	1.42	146.64	1.97	146.55	2.00	146.52
		60	1.56	146.45	1.98	146.42	2.00	146.55
	1000	40	1.20	146.80	1.72	146.49	1.97	146.51
		50	1.46	146.70	1.92	146.50	2.00	146.47
		60	1.53	146.56	1.99	146.56	2.00	146.51

Table 3.6: Standard errors for average true positives and average true negatives given in Table 3.5. The maximum number of true positives for a given replication is 2. The maximum number of true negatives for a given replication is $T - 3$

			δ=0.15		δ=0.25		δ=0.35	
T	p	n	ATP SE	ATN SE	ATP SE	ATN SE	ATP SE	ATN SE
50	500	40	0.40	0.62	0.47	0.52	0.17	0.49
		50	0.49	0.53	0.29	0.52	0.00	0.49
		60	0.50	0.55	0.14	0.50	0.00	0.55
	750	40	0.46	0.42	0.42	0.50	0.00	0.61
		50	0.47	0.48	0.22	0.50	0.00	0.46
		60	0.50	0.55	0.10	0.56	0.00	0.50
	1000	40	0.45	0.55	0.39	0.51	0.22	0.50
		50	0.50	0.47	0.22	0.50	0.00	0.43
		60	0.48	0.63	0.10	0.47	0.00	0.67
100	500	40	0.45	0.50	0.44	0.50	0.14	0.52
		50	0.47	0.47	0.27	0.50	0.00	0.50
		60	0.49	0.48	0.10	0.52	0.00	0.54
	750	40	0.42	0.50	0.36	0.51	0.14	0.50
		50	0.47	0.55	0.20	0.50	0.00	0.50
		60	0.49	0.49	0.10	0.50	0.00	0.78
	1000	40	0.40	0.43	0.44	0.50	0.14	0.55
		50	0.48	0.50	0.30	0.52	0.00	0.50
		60	0.49	0.61	0.00	0.50	0.00	0.61
150	500	40	0.39	0.43	0.45	0.56	0.17	0.56
		50	0.48	0.47	0.22	0.50	0.00	0.53
		60	0.50	0.52	0.00	0.50	0.00	0.52
	750	40	0.37	0.40	0.45	0.50	0.17	0.58
		50	0.50	0.50	0.17	0.52	0.00	0.56
		60	0.50	0.58	0.14	0.52	0.00	0.56
	1000	40	0.40	0.40	0.45	0.50	0.18	0.52
		50	0.50	0.46	0.27	0.67	0.00	0.63
		60	0.50	0.54	0.10	0.52	0.00	0.50

3.7 An empirical study

Human memory has been studied through fMRI experiments in the context of discrete and continuous activities. One goal of neurologists is to better understand perception and memory processes in humans as they experience continuous real-world events (Baldassano et al. 2017). Event segmentation theory, posited by Zacks et al. (2007), poses that under certain conditions, humans generate event boundaries in memory during continuous perception events. Thus, humans may partition a continuous experience into a series of segmented discrete events. Baldassano et al. (2017) investigated event boundary detection and concluded that long-term memory in humans is structured as a series of hierarchical discrete events. Moreover, Schapiro et al. (2013) suggested that event boundaries are formed around changes in functional connectivity. In this section, we apply our method to the task-based fMRI data set analyzed in Baldassano et al. (2017) and Chen et al. (2017) in order to study the brain's dynamic functional connectivity. In the presence of brain dynamic functional activity, points of change may represent these event boundaries as suggested in the aforementioned neuroscience literature.

We apply our proposed method to a task-based fMRI data set collected by Chen et al. (2017), where they investigated the effects of memories across different individuals. The experiment involved 17 participants that each watched the same 48-minute segment of the BBC television series Sherlock while undergoing an fMRI scan. The 48-minute segment was the first 48-minutes of the first episode in the television series. None of the participants had watched the series Sherlock prior to the study. Chen et al. partitioned the television episode into a 23-minute segment and a 25-minute segment. Each segment was prepended by a 30-second cartoon to allow the brain time to adjust to new audio and visual stimuli. Including an unrelated cartoon prior to studies such as this is common practice as it reduces statistical noise. Subjects were instructed to watch the television episode as they would watch a typical television episode in their own home. The fMRI data were gathered from a Siemens Skyra 3T full-body scanner. More details about the experiment and processes of acquiring functional and anatomical images are provided in Chen et al. (2017).

The 48-minute segment of Sherlock resulted in 1,976 time point measurements of data. For each participant, the fMRI machine acquired an image the participant's brain every 1.5 seconds. To demonstrate our proposed method, we analyzed the first 100 time points which equates to the first 150 seconds of the Sherlock episode. Let Y_{it} be the BOLD random vector for the 268 nodes of the ith individual at time t. Thus, Y_{it} $(i = 1, \ldots, 17;\ t = 1, \ldots, 100)$

is a 268-dimensional random vector. A node, or region of interest, represents a collection of voxels. The 268 node parcellation was performed according to Shen et al. (2013), where voxels groupings ensure functional homogeneity within each node, making it ideal for node network and dynamic functional connectivity analysis. Figure 3.2 illustrates the 268 Shen node parcellation along with large-scale node groupings. Node-level analysis decreases the data dimensional and allows for more interpretable results. For further details on the benefits and processes of Shen node parcellation, we refer readers to Shen et al. (2013).

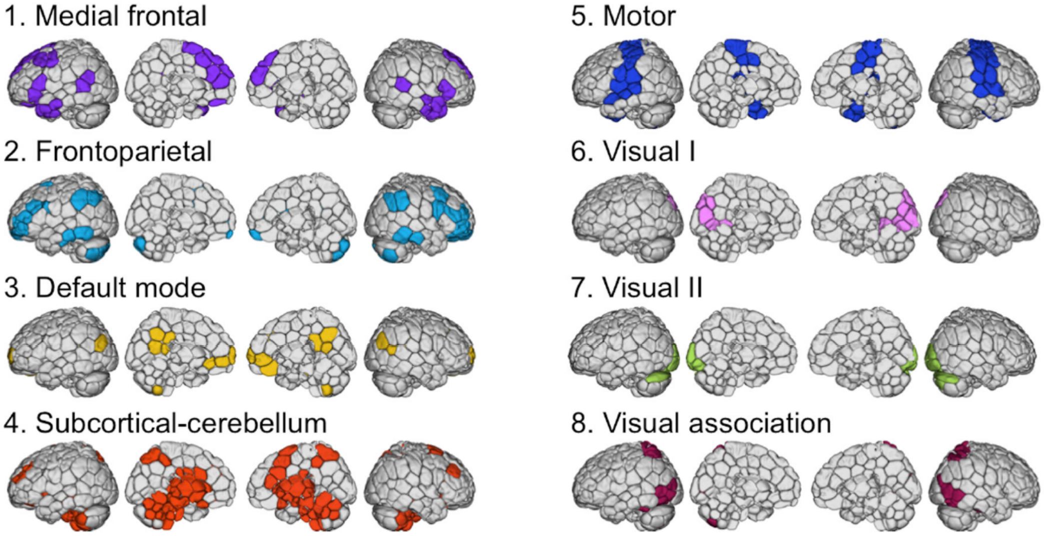

Figure 3.2: Shen 268 node parcellation. This image was obtained from Finn et al. (2015).

In our analysis $n = 17$, $p = 268$, and $T = 100$. Based on (3.1) – (3.2), we assume that at each time point there exists a common population covariance matrix among all 17 individuals. Our assumption is not unrealistic given this task-based fMRI experiment. In Chen et al. (2017) and Baldassano et al. (2017), they found that an across-subject design was appropriate due to consistent stimulus-response across patients for a given brain region.

Under model (3.2), we applied our procedure to test (3.1). Based on the test statistic value, $\mathcal{M}_n = 3.6596$, we rejected H_0 of (3.1) as the p-value was less than 0.001. Hence, we rejected the claim that the covariance matrices were stationary for all $T = 100$ time points. Accordingly, we applied binary segmentation to identify all significant change points among

99 possible points of change. Our proposed method identified 17 locations of significance. Change points were located at time points 2, 25, 36, 39, 40, 41, 42, 58, 60, 61, 63, 81, 83, 88, 89, 91, and 92. A change point at time two implies that $\Sigma_1 = \Sigma_2 \neq \Sigma_3$.

Table 3.7: Identified change points in the Sherlock fMRI data set. Range of time points preceding the identified change point where the covariance matrices are temporally homogeneous. An interval ID provides a reference to Figure 3.3

Change point	Interval	Homogeneous interval
2	1	[1, 2]
25	2	[3, 25]
36	3	[26, 36]
39	4	[37, 39]
40	5	[40, 40]
41	6	[41, 41]
42	7	[42, 42]
58	8	[43, 58]
60	9	[59, 60]
61	10	[61, 61]
63	11	[62, 63]
81	12	[64, 81]
83	13	[82, 83]
88	14	[84, 88]
89	15	[89, 89]
91	16	[90, 91]
92	17	[92, 92]

Figures (3.3) illustrates these temporal changes among covariance matrices around the identified change points listed in Table 3.7. Each subplot is the average correlation between nodes across the time interval where the covariance matrices are homogeneous. Thus, in Figure 3.3, Interval 1 represents the correlation network based on the average correlations between nodes over time interval $[1, 2]$. Interval 2 represents the correlation network based on the average correlations between nodes over time interval $[3, 25]$. Table 3.7 details the time interval corresponding to the temporal homogeneous covariance matrices preceding each identified change point. Therefore, given that a change point was located at $t = 2$, the correlation network of Interval 1 compared to Interval 2 should be significantly different. Correlation network layouts are structured according to the eight large-scale node groupings illustrated in Figure 3.2. The top-centered circle consists of nodes within the medial frontal group. Moving clockwise on a given sub-plot, the remaining circles represent frontoparietal, default mode, subcortical-cerebellum, motor, visual I, visual II, and visual association.

The identified change points in Table 3.7 coincide with interesting events in the television episode Sherlock. For example, the first change point at $t = 2$ may be a reaction to initial stimuli of the cartoon. The brain must process this initial video and audio stimuli. At approximately 37 to 38 seconds into the series Sherlock the cartoon ends, and a graphic war scene commences. Guns are fired, casualties are shown, but there is no distinguishable dialect. The transition point from cartoon clip to battle scene coincides with the change point identified at $t = 25$. After this war scene a period of quiet ensues. The first understandable dialect from actors occurs at approximately two minutes and 11 seconds into the episode. At this time, a therapist inquires about a patient's well-being as the viewer learns the opening war scene was a flashback. Change points identified at 88, 89, 91, and 92 equate to the start of this conversation.

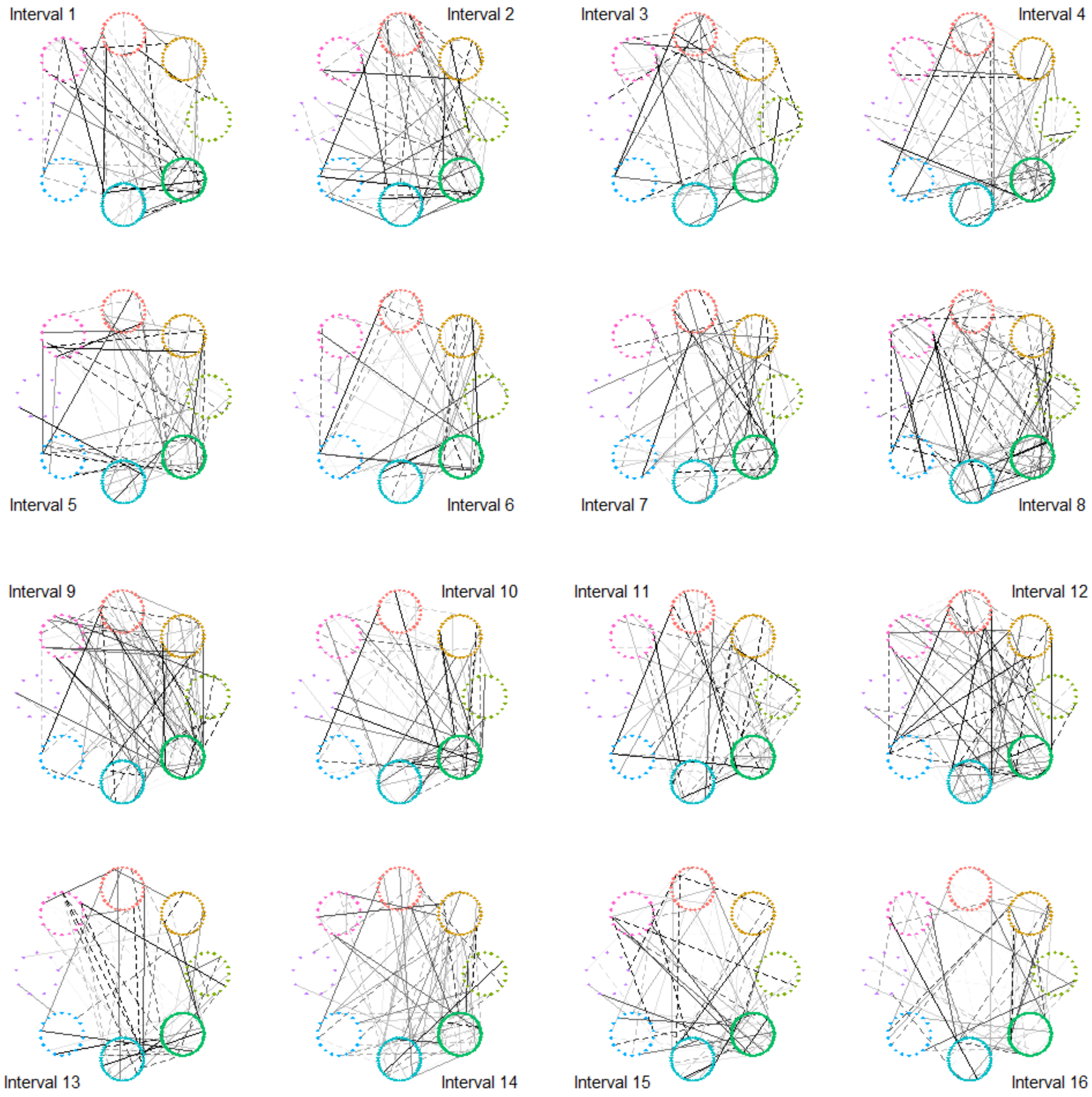

Figure 3.3: Correlation networks based on an average over a time interval in which the covariance matrices are homogeneous. Each circle is comprised of 67 Shen nodes. Solid lines represent a positive correlation, and dashed lines represent a negative correlation. The darker the line the stronger the correlation between nodes. A correlation threshold value of 0.70 in absolute values was used.

3.8 Technical details

This section contains proofs of lemmas and the main theorems. Some of the expressions are rather long. Thus, for readability, an equation will not always be aligned with the initial equality sign.

3.8.1 Proofs of lemmas

First, we provide proofs for some lemmas that will be used in the proofs corresponding to the main theorems.

Lemma 7. *Let* $Y_{is_ak} = \gamma_{s_ak}^{\mathrm{T}} z_i$*, where* z_i *is a standard multivariate normal random vector, and* $\gamma_{s_ak}^{\mathrm{T}}$ *is known. Then*

$$
\begin{aligned}
E\Big(Y_{is_ak}Y_{is_al}Y_{ir_cm}Y_{ir_cn}Y_{iu_eo}Y_{iu_ep}Y_{ix_gq}Y_{ix_gw}\Big) \asymp{}& \Sigma_{s_a}^{lk}\Sigma_{r_c}^{nm}\Sigma_{u_e}^{po}\Sigma_{x_g}^{wq} + \Sigma_{s_a}^{lk}\Sigma_{r_c}^{nm}C_{u_ex_g}^{pq}C_{x_gu_e}^{wo} \\
&+ \Sigma_{s_a}^{lk}\Sigma_{u_e}^{po}C_{r_cx_g}^{nq}C_{x_gr_c}^{wm} + \Sigma_{s_a}^{lk}\Sigma_{x_g}^{wq}C_{r_cu_e}^{no}C_{u_er_c}^{pm} + \Sigma_{r_c}^{nm}\Sigma_{u_e}^{po}C_{s_ax_g}^{lq}C_{x_gs_a}^{wk} \\
&+ \Sigma_{r_c}^{nm}\Sigma_{x_g}^{wq}C_{s_au_e}^{lo}C_{u_es_a}^{pk} + \Sigma_{u_e}^{po}\Sigma_{x_g}^{wq}C_{s_ar_c}^{lm}C_{r_cs_a}^{nk} + C_{s_ar_c}^{lm}C_{r_cs_a}^{nk}C_{u_ex_g}^{pq}C_{x_gu_e}^{wo} \\
&+ C_{s_au_e}^{lo}C_{u_es_a}^{pk}C_{r_cx_g}^{nq}C_{x_gr_c}^{wm} + C_{s_ax_g}^{lq}C_{x_gs_a}^{wk}C_{r_cu_e}^{no}C_{u_er_c}^{pm} + \Sigma_{s_a}^{lk}C_{r_cu_e}^{no}C_{u_ex_g}^{pq}C_{x_gr_c}^{wm} \\
&+ \Sigma_{r_c}^{nm}C_{s_au_e}^{lo}C_{u_ex_g}^{pq}C_{x_gs_a}^{wk} + \Sigma_{u_e}^{po}C_{s_ar_c}^{lm}C_{r_cx_g}^{nq}C_{x_gs_a}^{wk} + \Sigma_{x_g}^{wq}C_{s_ar_c}^{lm}C_{r_cu_e}^{no}C_{u_es_a}^{pk} \\
&+ C_{s_ar_c}^{lm}C_{r_cu_e}^{no}C_{u_ex_g}^{pq}C_{x_gs_a}^{wk} + C_{s_ar_c}^{lm}C_{r_cx_g}^{nq}C_{x_gu_e}^{wo}C_{u_es_a}^{pk} + C_{s_au_e}^{lo}C_{u_er_c}^{pm}C_{r_cx_g}^{nq}C_{x_gs_a}^{wk},
\end{aligned}
$$

where $\Sigma_{s_a}^{lk} = \gamma_{s_al}^{\mathrm{T}}\gamma_{s_ak}$ *and* $C_{u_ex_g}^{pq} = \gamma_{u_eq}^{\mathrm{T}}\gamma_{x_gq}$.

Proof. Let A_1, A_2, A_3, and A_4 be any matrices of appropriate dimensions. Assume z_i is a standard multivariate normal random vector. By the results of multivariate analysis

$$
\begin{aligned}
E\Big(z_i^T A_1 z_i z_i^T A_2 z_i z_i^T A_3 z_i z_i^T A_4 z_i\Big) &= \mathrm{tr}(A_1)\mathrm{tr}(A_2)\mathrm{tr}(A_3)\mathrm{tr}(A_4) \\
&+ \Big[\mathrm{tr}(A_1)\mathrm{tr}(A_2)\mathrm{tr}(A_3A_4 + A_3A_4^T) + \mathrm{tr}(A_1)\mathrm{tr}(A_3)\mathrm{tr}(A_2A_4 + A_2A_4^T) \\
&+ \mathrm{tr}(A_1)\mathrm{tr}(A_4)\mathrm{tr}(A_2A_3 + A_2A_3^T) + \mathrm{tr}(A_2)\mathrm{tr}(A_3)\mathrm{tr}(A_1A_4 + A_1A_4^T) \\
&+ \mathrm{tr}(A_2)\mathrm{tr}(A_4)\mathrm{tr}(A_1A_3 + A_1A_3^T) + \mathrm{tr}(A_3)\mathrm{tr}(A_4)\mathrm{tr}(A_1A_2 + A_1A_2^T)\Big] \\
&+ \Big[\mathrm{tr}(A_1A_2 + A_1A_2^T)\mathrm{tr}(A_3A_4 + A_3A_4^T) + \mathrm{tr}(A_1A_3 + A_1A_3^T)\mathrm{tr}(A_2A_4 + A_2A_4^T) \\
&+ \mathrm{tr}(A_1A_4 + A_1A_4^T)\mathrm{tr}(A_2A_3 + A_2A_3^T)\Big] + \Big[\mathrm{tr}(A_1)\mathrm{tr}\{(A_2 + A_2^T)(A_3 + A_3^T)(A_4 + A_4^T)\} \\
&+ \mathrm{tr}(A_2)\mathrm{tr}\{(A_1 + A_1^T)(A_3 + A_3^T)(A_4 + A_4^T)\} \\
&+ \mathrm{tr}(A_3)\mathrm{tr}\{(A_1 + A_1^T)(A_2 + A_2^T)(A_4 + A_4^T)\} \\
&+ \mathrm{tr}(A_4)\mathrm{tr}\{(A_1 + A_1^T)(A_2 + A_2^T)(A_3 + A_3^T)\}\Big] \\
&+ \Big[\mathrm{tr}\{(A_1 + A_1^T)(A_2 + A_2^T)(A_3 + A_3^T)(A_4 + A_4^T)\} \\
&+ \mathrm{tr}\{(A_1 + A_1^T)(A_2 + A_2^T)(A_4 + A_4^T)(A_3 + A_3^T)\} \\
&+ \mathrm{tr}\{(A_1 + A_1^T)(A_3 + A_3^T)(A_2 + A_2^T)(A_4 + A_4^T)\}\Big].
\end{aligned}
$$

By the definition of $Y_{i..}$, $E\Big\{Y_{is_ak}Y_{is_al}Y_{ir_cm}Y_{ir_cn}Y_{iu_eo}Y_{iu_ep}Y_{ix_gq}Y_{ix_gw}\Big\} =$ $E\Big\{z_i^T\gamma_{s_ak}\gamma_{s_al}^T z_i z_i^T\gamma_{r_cm}\gamma_{r_cn}^T z_i z_i^T\gamma_{u_eo}\gamma_{u_ep}^T z_i z_i^T\gamma_{x_gq}\gamma_{x_gw}^T z_i\Big\}$. Thus, making the appropriate substitutions for A_1, A_2, A_3, and A_4, it follows that

$$
\begin{aligned}
E\Big(Y_{is_ak}Y_{is_al}Y_{ir_cm}Y_{ir_cn}Y_{iu_eo}Y_{iu_ep}Y_{ix_gq}Y_{ix_gw}\Big) &\asymp \Sigma_{sa}^{lk}\Sigma_{rc}^{nm}\Sigma_{ue}^{po}\Sigma_{xg}^{wq} + \Sigma_{sa}^{lk}\Sigma_{rc}^{nm}C_{uexg}^{pq}C_{xgue}^{wo} \\
&+ \Sigma_{sa}^{lk}\Sigma_{ue}^{po}C_{rcxg}^{nq}C_{xgrc}^{wm} + \Sigma_{sa}^{lk}\Sigma_{xg}^{wq}C_{rcue}^{no}C_{uerc}^{pm} + \Sigma_{rc}^{nm}\Sigma_{ue}^{po}C_{saxg}^{lq}C_{xgsa}^{wk} \\
&+ \Sigma_{rc}^{nm}\Sigma_{xg}^{wq}C_{saue}^{lo}C_{uesa}^{pk} + \Sigma_{ue}^{po}\Sigma_{xg}^{wq}C_{sarc}^{lm}C_{rcsa}^{nk} + C_{sarc}^{lm}C_{rcsa}^{nk}C_{uexg}^{pq}C_{xgue}^{wo} \\
&+ C_{saue}^{lo}C_{uesa}^{pk}C_{rcxg}^{nq}C_{xgrc}^{wm} + C_{saxg}^{lq}C_{xgsa}^{wk}C_{rcue}^{no}C_{uerc}^{pm} + \Sigma_{sa}^{lk}C_{rcue}^{no}C_{uexg}^{pq}C_{xgrc}^{wm} \\
&+ \Sigma_{rc}^{nm}C_{saue}^{lo}C_{uexg}^{pq}C_{xgsa}^{wk} + \Sigma_{ue}^{po}C_{sarc}^{lm}C_{rcxg}^{nq}C_{xgsa}^{wk} + \Sigma_{xg}^{wq}C_{sarc}^{lm}C_{rcue}^{no}C_{uesa}^{pk} \\
&+ C_{sarc}^{lm}C_{rcue}^{no}C_{uexg}^{pq}C_{xgsa}^{wk} + C_{sarc}^{lm}C_{rcxg}^{nq}C_{xgue}^{wo}C_{uesa}^{pk} + C_{saue}^{lo}C_{uerc}^{pm}C_{rcxg}^{nq}C_{xgsa}^{wk},
\end{aligned}
$$

where $\Sigma_{sa}^{lk} = \gamma_{s_al}^{\mathrm{T}}\gamma_{s_ak}$ and $C_{uexg}^{pq} = \gamma_{u_eq}^{\mathrm{T}}\gamma_{x_gq}$. $\square$

Lemma 8. *Let* $V_{s_as_b}(i,j) = (Y_{is_a}^T Y_{js_b})^2$*, where* $s_a, s_b, r_c, r_d, u_e, u_f, x_g, x_h \in \{1 \ldots T-1\}$ *and* $i \neq j$*. Then*

$$\begin{aligned} E\Big\{V_{s_as_b}(i,j)V_{r_cr_d}(i,j)V_{u_eu_f}(i,j)V_{x_gx_h}(i,j)\Big\} = {} & \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3 + \mathcal{H}_4 + \mathcal{H}_5 + \mathcal{H}_6 \\ & + \mathcal{H}_7 + \mathcal{H}_8 + \mathcal{H}_9 + \mathcal{H}_{10} + \mathcal{H}_{11} + \mathcal{H}_{12} \\ & + \mathcal{H}_{13} + \mathcal{H}_{14} + \mathcal{H}_{15} + \mathcal{H}_{16} + \mathcal{H}_{17} \end{aligned}$$

where $\mathcal{H}_k, k \in \{1, \ldots, 17\}$ *is given below.*

$$\begin{aligned} \mathcal{H}_1 = \Big[& \mathrm{tr}^4(\Sigma^2) + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{u_fx_h}\Sigma C_{x_hu_f}) + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{r_dx_h}\Sigma C_{x_hr_d}) \\ & + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{r_du_f}\Sigma C_{u_fr_d}) + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bx_h}\Sigma C_{x_hs_b}) \\ & + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bu_f}\Sigma C_{u_fs_b}) + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{s_br_d}\Sigma C_{r_ds_b}) \\ & + \mathrm{tr}(\Sigma C_{s_br_d}\Sigma C_{r_ds_b})\mathrm{tr}(\Sigma C_{u_fx_h}\Sigma C_{x_hu_f}) + \mathrm{tr}(\Sigma C_{s_bu_f}\Sigma C_{u_fs_b})\mathrm{tr}(\Sigma C_{r_dx_h}\Sigma C_{x_hr_d}) \\ & + \mathrm{tr}(\Sigma C_{s_bx_h}\Sigma C_{x_hs_b})\mathrm{tr}(\Sigma C_{r_du_f}\Sigma C_{u_fr_d}) + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{r_dx_h}\Sigma C_{x_hu_f}\Sigma C_{u_fr_d}) \\ & + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bx_h}\Sigma C_{x_hu_f}\Sigma C_{u_fs_b}) + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bx_h}\Sigma C_{x_hr_d}\Sigma C_{r_ds_b}) \\ & + +\mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bu_f}\Sigma C_{u_fr_d}\Sigma C_{r_ds_b}) + \mathrm{tr}(\Sigma C_{s_bx_h}\Sigma C_{x_hu_f}\Sigma C_{u_fr_d}\Sigma C_{r_ds_b}) \\ & + \mathrm{tr}(\Sigma C_{s_bu_f}\Sigma C_{u_fx_h}\Sigma C_{x_hr_d}\Sigma C_{r_ds_b}) + \mathrm{tr}(\Sigma C_{s_bx_h}\Sigma C_{x_hr_d}\Sigma C_{r_du_f}\Sigma C_{u_fs_b})\Big] \end{aligned}$$

$$\begin{aligned} \mathcal{H}_2 = \Big[& \mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{u_ex_g}\Sigma C_{x_gu_e}) + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}^2(C_{u_ex_g}C_{x_hu_f}) \\ & + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{r_dx_h}C_{x_gu_e}\Sigma C_{u_ex_g}C_{x_hr_d}) \\ & + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{r_du_f}C_{u_ex_g}\Sigma C_{x_gu_e}C_{u_fr_d}) + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bx_h}C_{x_gu_e}\Sigma C_{u_ex_g}C_{x_hs_b}) \\ & + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bu_f}C_{u_ex_g}\Sigma C_{x_gu_e}C_{u_fs_b}) + \mathrm{tr}(\Sigma C_{s_br_d}\Sigma C_{r_ds_b})\mathrm{tr}(\Sigma C_{u_ex_g}\Sigma C_{x_gu_e}) \\ & + \mathrm{tr}(\Sigma C_{s_br_d}\Sigma C_{r_ds_b})\mathrm{tr}^2(C_{u_ex_g}C_{x_hu_f}) \\ & + \mathrm{tr}(\Sigma C_{s_bu_f}C_{u_ex_g}C_{x_hr_d}\Sigma C_{r_dx_h}C_{x_gu_e}C_{u_fs_b}) \\ & + \mathrm{tr}(\Sigma C_{s_bx_h}C_{x_gu_e}C_{u_fr_d}\Sigma C_{r_du_f}C_{u_ex_g}C_{x_hs_b}) \\ & + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{r_dx_h}C_{x_gu_e}C_{u_fr_d})\mathrm{tr}(C_{u_ex_g}C_{x_hu_f}) \end{aligned}$$

$$
\begin{aligned}
&+ \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g u_e} C_{u_f s_b})\operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g u_e} \Sigma C_{u_e x_g} C_{x_h r_d} \Sigma C_{r_d s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b u_f} C_{u_e x_g} \Sigma C_{x_g u_e} C_{u_f r_d} \Sigma C_{r_d s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g u_e} C_{r_d u_f} \Sigma C_{r_d s_b})\operatorname{tr}^2(C_{u_e x_g} C_{x_h u_f}) \\
&+ \operatorname{tr}(\Sigma C_{s_b u_f} C_{u_e x_g} C_{x_h r_d} \Sigma C_{r_d s_b})\operatorname{tr}^2(C_{u_e x_g} C_{x_h u_f}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g u_e} C_{u_f s_b})\operatorname{tr}(\Sigma C_{r_d u_f} C_{u_e x_g} C_{x_h r_d})\Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_3 = \Big[&\operatorname{tr}^2(\Sigma^2)\operatorname{tr}(\Sigma C_{r_c x_g} \Sigma C_{x_g r_c}) + \operatorname{tr}^2(\Sigma^2)\operatorname{tr}^2(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g r_c} \Sigma C_{r_c x_g} C_{x_h u_f}) \\
&+ \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{u_f r_d} C_{r_c x_g} \Sigma C_{x_g r_c} C_{r_d u_f}) + \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} \Sigma C_{r_c x_g} C_{x_h s_b}) \\
&+ \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{s_b r_d} C_{r_c x_g} \Sigma C_{x_g r_c} C_{r_d s_b}) + \operatorname{tr}(\Sigma C_{s_b u_f} \Sigma C_{u_f s_b})\operatorname{tr}(\Sigma C_{r_c x_g} \Sigma C_{x_g r_c}) \\
&+ \operatorname{tr}(\Sigma C_{s_b u_f} \Sigma C_{u_f s_b})\operatorname{tr}^2(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma C_{s_b r_d} C_{r_c x_g} C_{x_h u_f} \Sigma C_{u_f x_h} C_{x_g r_c} C_{r_d s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{r_d u_f} \Sigma C_{u_f r_d} C_{r_c x_g} C_{x_h s_b}) \\
&+ \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g r_c} C_{r_d u_f})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{r_d s_b})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} \Sigma C_{r_c x_g} C_{x_h u_f} \Sigma C_{u_f s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b r_d} C_{r_c x_g} \Sigma C_{x_g r_c} C_{r_d u_f} \Sigma C_{u_f s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{u_f r_d} \Sigma C_{u_f s_b})\operatorname{tr}^2(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma C_{s_b r_d} C_{r_c x_g} C_{x_h u_f} \Sigma C_{u_f s_b})\operatorname{tr}^2(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{r_d s_b})\operatorname{tr}(\Sigma C_{u_f r_d} C_{r_c x_g} C_{x_h u_f})\Big]
\end{aligned}
$$

$$
\mathcal{H}_4 = \Big[\operatorname{tr}^2(\Sigma^2)\operatorname{tr}(\Sigma C_{r_c u_e} \Sigma C_{u_e r_c}) + \operatorname{tr}^2(\Sigma^2)\operatorname{tr}^2(C_{r_c u_e} C_{u_f r_d})
$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{x_h u_f} C_{u_e r_c} \Sigma C_{r_c u_e} C_{u_f x_h})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{x_h r_d} C_{r_c u_e} \Sigma C_{u_e r_c} C_{r_d x_h}) + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_b u_f} C_{u_e r_c} \Sigma C_{r_c u_e} C_{u_f s_b})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_b r_d} C_{r_c u_e} \Sigma C_{u_e r_c} C_{r_d s_b}) + \mathrm{tr}(\Sigma C_{s_b x_h} \Sigma C_{x_h s_b})\mathrm{tr}(\Sigma C_{r_c u_e} \Sigma C_{u_e r_c})$$

$$+ \mathrm{tr}(\Sigma C_{s_b x_h} \Sigma C_{x_h s_b})\mathrm{tr}^2(C_{r_c u_e} C_{u_f r_d}) + \mathrm{tr}(\Sigma C_{s_b r_d} C_{r_c u_e} C_{u_f x_h} \Sigma C_{x_h u_f} C_{u_e r_c} C_{r_d s_b})$$

$$+ \mathrm{tr}(\Sigma C_{s_b u_f} C_{u_e r_c} C_{r_d x_h} \Sigma C_{x_h r_d} C_{r_c u_e} C_{u_f s_b})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{x_h u_f} C_{u_e r_c} C_{r_d x_h})\mathrm{tr}(C_{r_c u_e} C_{u_f r_d})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{s_b u_f} C_{u_e r_c} C_{r_d s_b})\mathrm{tr}(C_{r_c u_e} C_{u_f r_d})$$

$$+ \mathrm{tr}(\Sigma C_{s_b u_f} C_{u_e r_c} \Sigma C_{r_c u_e} C_{u_f x_h} \Sigma C_{x_h s_b})$$

$$+ \mathrm{tr}(\Sigma C_{s_b r_d} C_{r_c u_e} \Sigma C_{u_e r_c} C_{r_d x_h} \Sigma C_{x_h s_b})$$

$$+ \mathrm{tr}(\Sigma C_{s_b u_f} C_{u_e r_c} C_{x_h r_d} \Sigma C_{x_h s_b})\mathrm{tr}^2(C_{r_c u_e} C_{u_f r_d})$$

$$+ \mathrm{tr}(\Sigma C_{s_b r_d} C_{r_c u_e} C_{u_f x_h} \Sigma C_{x_h s_b})\mathrm{tr}^2(C_{r_c u_e} C_{u_f r_d})$$

$$+ \mathrm{tr}(\Sigma C_{s_b u_f} C_{u_e r_c} C_{r_d s_b})\mathrm{tr}(\Sigma C_{x_h r_d} C_{r_c u_e} C_{u_f x_h})\Big]$$

$$\mathcal{H}_5 = \Big[\mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g s_a}) + \mathrm{tr}^2(\Sigma^2)\mathrm{tr}^2(C_{s_a x_g} C_{x_h s_b})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} \Sigma C_{s_a x_g} C_{x_h u_f})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{u_f s_b} C_{s_a x_g} \Sigma C_{x_g s_a} C_{s_b u_f}) + \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} \Sigma C_{s_a x_g} C_{x_h r_d})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a x_g} \Sigma C_{x_g s_a} C_{s_b r_d}) + \mathrm{tr}(\Sigma C_{r_d u_f} \Sigma C_{u_f r_d})\mathrm{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g s_a})$$

$$+ \mathrm{tr}(\Sigma C_{r_d u_f} \Sigma C_{u_f r_d})\mathrm{tr}^2(C_{s_a x_g} C_{x_h s_b})$$

$$+ \mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a x_g} C_{x_h u_f} \Sigma C_{u_f x_h} C_{x_g s_a} C_{s_b r_d})$$

$$+ \mathrm{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{s_b u_f} \Sigma C_{u_f s_b} C_{s_a x_g} C_{x_h r_d})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} C_{s_b u_f})\mathrm{tr}(C_{s_a x_g} C_{x_h s_b})$$

$$+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{s_b r_d})\mathrm{tr}(C_{s_a x_g} C_{x_h s_b})$$

$$+ \mathrm{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} \Sigma C_{s_a x_g} C_{x_h u_f} \Sigma C_{u_f r_d})$$

$$
\begin{aligned}
&+ \mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a x_g} \Sigma C_{x_g s_a} C_{s_b u_f} \Sigma C_{u_f r_d}) \\
&+ \mathrm{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{u_f s_b} \Sigma C_{u_f r_d}) \mathrm{tr}^2(C_{s_a x_g} C_{x_h s_b}) \\
&+ \mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a x_g} C_{x_h u_f} \Sigma C_{u_f r_d}) \mathrm{tr}^2(C_{s_a x_g} C_{x_h s_b}) \\
&+ \mathrm{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{s_b r_d}) \mathrm{tr}(\Sigma C_{u_f s_b} C_{s_a x_g} C_{x_h u_f}) \Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_6 = \Big[&\mathrm{tr}^2(\Sigma^2) \mathrm{tr}(\Sigma C_{s_a u_e} \Sigma C_{u_e s_a}) + \mathrm{tr}^2(\Sigma^2) \mathrm{tr}^2(C_{s_a u_e} C_{u_f s_b}) \\
&+ \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{x_h u_f} C_{u_e s_a} \Sigma C_{s_a u_e} C_{u_f x_h}) \\
&+ \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{x_h s_b} C_{s_a u_e} \Sigma C_{u_e s_a} C_{s_b x_h}) + \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{r_d u_f} C_{u_e s_a} \Sigma C_{s_a u_e} C_{u_f r_d}) \\
&+ \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a u_e} \Sigma C_{u_e s_a} C_{s_b r_d}) + \mathrm{tr}(\Sigma C_{r_d x_h} \Sigma C_{x_h r_d}) \mathrm{tr}(\Sigma C_{s_a u_e} \Sigma C_{u_e s_a}) \\
&+ \mathrm{tr}(\Sigma C_{r_d x_h} \Sigma C_{x_h r_d}) \mathrm{tr}^2(C_{s_a u_e} C_{u_f s_b}) \\
&+ \mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a u_e} C_{u_f x_h} \Sigma C_{x_h u_f} C_{u_e s_a} C_{s_b r_d}) \\
&+ \mathrm{tr}(\Sigma C_{r_d u_f} C_{u_e s_a} C_{s_b x_h} \Sigma C_{x_h s_b} C_{s_a u_e} C_{u_f r_d}) \\
&+ \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{x_h u_f} C_{u_e s_a} C_{s_b x_h}) \mathrm{tr}(C_{s_a u_e} C_{u_f s_b}) \\
&+ \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{r_d u_f} C_{u_e s_a} C_{s_b r_d}) \mathrm{tr}(C_{s_a u_e} C_{u_f s_b}) \\
&+ \mathrm{tr}(\Sigma C_{r_d u_f} C_{u_e s_a} \Sigma C_{s_a u_e} C_{u_f x_h} \Sigma C_{x_h r_d}) \\
&+ \mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a u_e} \Sigma C_{u_e s_a} C_{s_b x_h} \Sigma C_{x_h r_d}) \\
&+ \mathrm{tr}(\Sigma C_{r_d u_f} C_{u_e s_a} C_{x_h s_b} \Sigma C_{x_h r_d}) \mathrm{tr}^2(C_{s_a u_e} C_{u_f s_b}) \\
&+ \mathrm{tr}(\Sigma C_{r_d s_b} C_{s_a u_e} C_{u_f x_h} \Sigma C_{x_h r_d}) \mathrm{tr}^2(C_{s_a u_e} C_{u_f s_b}) \\
&+ \mathrm{tr}(\Sigma C_{r_d u_f} C_{u_e s_a} C_{s_b r_d}) \mathrm{tr}(\Sigma C_{x_h s_b} C_{s_a u_e} C_{u_f x_h}) \Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_7 = \Big[&\mathrm{tr}^2(\Sigma^2) \mathrm{tr}(\Sigma C_{s_a r_c} \Sigma C_{r_c s_a}) + \mathrm{tr}^2(\Sigma^2) \mathrm{tr}^2(C_{s_a r_c} C_{r_d s_b}) \\
&+ \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{x_h r_d} C_{r_c s_a} \Sigma C_{s_a r_c} C_{r_d x_h}) \\
&+ \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{x_h s_b} C_{s_a r_c} \Sigma C_{r_c s_a} C_{s_b x_h}) + \mathrm{tr}(\Sigma^2) \mathrm{tr}(\Sigma C_{u_f r_d} C_{r_c s_a} \Sigma C_{s_a r_c} C_{r_d u_f})
\end{aligned}
$$

$$
\begin{aligned}
&+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{u_f s_b} C_{s_a r_c} \Sigma C_{r_c s_a} C_{s_b u_f}) + \mathrm{tr}(\Sigma C_{u_f x_h} \Sigma C_{x_h u_f})\mathrm{tr}(\Sigma C_{s_a r_c} \Sigma C_{r_c s_a}) \\
&+ \mathrm{tr}(\Sigma C_{u_f x_h} \Sigma C_{x_h u_f})\mathrm{tr}^2(C_{s_a r_c} C_{r_d s_b}) + \mathrm{tr}(\Sigma C_{u_f s_b} C_{s_a r_c} C_{r_d x_h} \Sigma C_{x_h r_d} C_{r_c s_a} C_{s_b u_f}) \\
&+ \mathrm{tr}(\Sigma C_{u_f r_d} C_{r_c s_a} C_{s_b x_h} \Sigma C_{x_h s_b} C_{s_a r_c} C_{r_d u_f}) \\
&+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{x_h r_d} C_{r_c s_a} C_{s_b x_h})\mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \\
&+ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma C_{u_f r_d} C_{r_c s_a} C_{s_b u_f})\mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \\
&+ \mathrm{tr}(\Sigma C_{u_f r_d} C_{r_c s_a} \Sigma C_{s_a r_c} C_{r_d x_h} \Sigma C_{x_h u_f}) \\
&+ \mathrm{tr}(\Sigma C_{u_f s_b} C_{s_a r_c} \Sigma C_{r_c s_a} C_{s_b x_h} \Sigma C_{x_h u_f}) \\
&+ \mathrm{tr}(\Sigma C_{u_f r_d} C_{r_c s_a} C_{x_h s_b} \Sigma C_{x_h u_f})\mathrm{tr}^2(C_{s_a r_c} C_{r_d s_b}) \\
&+ \mathrm{tr}(\Sigma C_{u_f s_b} C_{s_a r_c} C_{r_d x_h} \Sigma C_{x_h u_f})\mathrm{tr}^2(C_{s_a r_c} C_{r_d s_b}) \\
&+ \mathrm{tr}(\Sigma C_{u_f r_d} C_{r_c s_a} C_{s_b u_f})\mathrm{tr}(\Sigma C_{x_h s_b} C_{s_a r_c} C_{r_d x_h})\Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_8 = \Big[&\mathrm{tr}(\Sigma C_{s_a r_c} \Sigma C_{r_c s_a})\mathrm{tr}(\Sigma C_{u_e x_g} \Sigma C_{x_g u_e}) + \mathrm{tr}(\Sigma C_{s_a r_c} \Sigma C_{r_c s_a})\mathrm{tr}^2(C_{u_e x_g} C_{x_h u_f}) \\
&+ \mathrm{tr}(\Sigma C_{s_a r_c} C_{r_d x_h} C_{x_g u_e} \Sigma C_{u_e x_g} C_{x_h r_d} C_{r_c s_a}) \\
&+ \mathrm{tr}(\Sigma C_{s_a r_c} C_{r_d u_f} C_{u_e x_g} \Sigma C_{x_g u_e} C_{u_f r_d} C_{r_c s_a}) \\
&+ \mathrm{tr}(\Sigma C_{r_c s_a} C_{s_b x_h} C_{x_g u_e} \Sigma C_{u_e x_g} C_{x_h s_b} C_{s_a r_c}) \\
&+ \mathrm{tr}(\Sigma C_{r_c s_a} C_{s_b u_f} C_{u_e x_g} \Sigma C_{x_g u_e} C_{u_f s_b} C_{s_a r_c}) \\
&+ \mathrm{tr}(\Sigma C_{u_e x_g} \Sigma C_{x_g u_e})\mathrm{tr}^2(C_{s_a r_c} C_{r_d s_b}) + \mathrm{tr}^2(C_{s_a r_c} C_{r_d s_b})\mathrm{tr}^2(C_{u_e x_g} C_{x_h u_f}) \\
&+ \mathrm{tr}^2(C_{s_a r_c} C_{r_d x_h} C_{x_g u_e} C_{u_f s_b}) + \mathrm{tr}^2(C_{s_a r_c} C_{r_d u_f} C_{u_e x_g} C_{x_h s_b}) \\
&+ \mathrm{tr}(\Sigma C_{s_a r_c} C_{r_d u_f} C_{u_e x_g} C_{x_h r_d} C_{r_c s_a})\mathrm{tr}(C_{u_e x_g} C_{x_h u_f}) \\
&+ \mathrm{tr}(\Sigma C_{r_c s_a} C_{s_b u_f} C_{u_e x_g} C_{x_h s_b} C_{s_a r_c})\mathrm{tr}(C_{u_e x_g} C_{x_h u_f}) \\
&+ \mathrm{tr}(\Sigma C_{u_e x_g} C_{x_h s_b} C_{s_a r_c} C_{r_d x_h} C_{x_g u_e})\mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \\
&+ \mathrm{tr}(\Sigma C_{x_g u_e} C_{u_f s_b} C_{s_a r_c} C_{r_d u_f} C_{u_e x_g})\mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \\
&+ \mathrm{tr}(C_{r_c s_a} C_{s_b x_h} C_{x_g u_e} C_{u_f r_d})\mathrm{tr}(C_{s_a r_c} C_{r_d s_b})\mathrm{tr}(C_{u_e x_g} C_{x_h u_f})
\end{aligned}
$$

$$
\begin{aligned}
&+\operatorname{tr}(C_{r_c s_a} C_{s_b u_f} C_{u_e x_g} C_{x_h r_d})\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})\operatorname{tr}(C_{u_e x_g} C_{x_h u_f})\\
&+\operatorname{tr}(C_{s_a r_c} C_{r_d u_f} C_{u_e x_g} C_{x_h r_d} C_{r_c s_a} C_{s_b x_h} C_{x_g u_e} C_{u_f s_b})\Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_9 = \Big[&\operatorname{tr}(\Sigma C_{s_a u_e} \Sigma C_{u_e s_a})\operatorname{tr}(\Sigma C_{r_c x_g} \Sigma C_{x_g r_c}) + \operatorname{tr}(\Sigma C_{s_a u_e} \Sigma C_{u_e s_a})\operatorname{tr}^2(C_{r_c x_g} C_{x_h r_d})\\
&+\operatorname{tr}(\Sigma C_{s_a u_e} C_{u_f x_h} C_{x_g r_c} \Sigma C_{r_c x_g} C_{x_h u_f} C_{u_e s_a})\\
&+\operatorname{tr}(\Sigma C_{s_a u_e} C_{u_f r_d} C_{r_c x_g} \Sigma C_{x_g r_c} C_{r_d u_f} C_{u_e s_a})\\
&+\operatorname{tr}(\Sigma C_{u_e s_a} C_{s_b x_h} C_{x_g r_c} \Sigma C_{r_c x_g} C_{x_h s_b} C_{s_a u_e})\\
&+\operatorname{tr}(\Sigma C_{u_e s_a} C_{s_b r_d} C_{r_c x_g} \Sigma C_{x_g r_c} C_{r_d s_b} C_{s_a u_e})\\
&+\operatorname{tr}(\Sigma C_{r_c x_g} \Sigma C_{x_g r_c})\operatorname{tr}^2(C_{s_a u_e} C_{u_f s_b}) + \operatorname{tr}^2(C_{s_a u_e} C_{u_f s_b})\operatorname{tr}^2(C_{r_c x_g} C_{x_h r_d})\\
&+\operatorname{tr}^2(C_{s_a u_e} C_{u_f x_h} C_{x_g r_c} C_{r_d s_b}) + \operatorname{tr}^2(C_{s_a u_e} C_{u_f r_d} C_{r_c x_g} C_{x_h s_b})\\
&+\operatorname{tr}(\Sigma C_{s_a u_e} C_{u_f r_d} C_{r_c x_g} C_{x_h u_f} C_{u_e s_a})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})\\
&+\operatorname{tr}(\Sigma C_{u_e s_a} C_{s_b r_d} C_{r_c x_g} C_{x_h s_b} C_{s_a u_e})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})\\
&+\operatorname{tr}(\Sigma C_{r_c x_g} C_{x_h s_b} C_{s_a u_e} C_{u_f x_h} C_{x_g r_c})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\\
&+\operatorname{tr}(\Sigma C_{x_g r_c} C_{r_d s_b} C_{s_a u_e} C_{u_f r_d} C_{r_c x_g})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\\
&+\operatorname{tr}(C_{u_e s_a} C_{s_b x_h} C_{x_g r_c} C_{r_d u_f})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})\\
&+\operatorname{tr}(C_{u_e s_a} C_{s_b r_d} C_{r_c x_g} C_{x_h u_f})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})\\
&+\operatorname{tr}(C_{s_a u_e} C_{u_f r_d} C_{r_c x_g} C_{x_h u_f} C_{u_e s_a} C_{s_b x_h} C_{x_g r_c} C_{r_d s_b})\Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_{10} = \Big[&\operatorname{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g s_a})\operatorname{tr}(\Sigma C_{r_c u_e} \Sigma C_{u_e r_c}) + \operatorname{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g s_a})\operatorname{tr}^2(C_{r_c u_e} C_{u_f r_d})\\
&+\operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h u_f} C_{u_e r_c} \Sigma C_{r_c u_e} C_{u_f x_h} C_{x_g s_a})\\
&+\operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h r_d} C_{r_c u_e} \Sigma C_{u_e r_c} C_{r_d x_h} C_{x_g s_a})\\
&+\operatorname{tr}(\Sigma C_{x_g s_a} C_{s_b u_f} C_{u_e r_c} \Sigma C_{r_c u_e} C_{u_f s_b} C_{s_a x_g})\\
&+\operatorname{tr}(\Sigma C_{x_g s_a} C_{s_b r_d} C_{r_c u_e} \Sigma C_{u_e r_c} C_{r_d s_b} C_{s_a x_g})
\end{aligned}
$$

$$
\begin{aligned}
&+ \operatorname{tr}(\Sigma C_{r_c u_e} \Sigma C_{u_e r_c}) \operatorname{tr}^2(C_{s_a x_g} C_{x_h s_b}) \\
&+ \operatorname{tr}^2(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}^2(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}^2(C_{s_a x_g} C_{x_h u_f} C_{u_e r_c} C_{r_d s_b}) + \operatorname{tr}^2(C_{s_a x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f x_h} C_{x_g s_a}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}(\Sigma C_{x_g s_a} C_{s_b r_d} C_{r_c u_e} C_{u_f s_b} C_{s_a x_g}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}(\Sigma C_{r_c u_e} C_{u_f s_b} C_{s_a x_g} C_{x_h u_f} C_{u_e r_c}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \\
&+ \operatorname{tr}(\Sigma C_{u_e r_c} C_{r_d s_b} C_{s_a x_g} C_{x_h r_d} C_{r_c u_e}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \\
&+ \operatorname{tr}(C_{x_g s_a} C_{s_b u_f} C_{u_e r_c} C_{r_d x_h}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}(C_{x_g s_a} C_{s_b r_d} C_{r_c u_e} C_{u_f x_h}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}(C_{s_a x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f x_h} C_{x_g s_a} C_{s_b u_f} C_{u_e r_c} C_{r_d s_b}) \Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_{11} = \Big[&\operatorname{tr}(\Sigma^2) \operatorname{tr}(\Sigma C_{r_c x_g} \Sigma C_{x_g u_e} \Sigma C_{u_e r_c}) + \operatorname{tr}(\Sigma^2) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \operatorname{tr}(\Sigma C_{r_c x_g} C_{x_h u_f} C_{u_e r_c}) \\
&+ \operatorname{tr}(\Sigma^2) \operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \operatorname{tr}(\Sigma C_{u_e r_c} C_{r_d x_h} C_{x_g u_e}) \\
&+ \operatorname{tr}(\Sigma^2) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \operatorname{tr}(\Sigma C_{x_g u_e} C_{u_f r_d} C_{r_c x_g}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} \Sigma C_{r_c u_e} \Sigma C_{u_e x_g} C_{x_h s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b u_f} C_{u_e x_g} \Sigma C_{x_g r_c} \Sigma C_{r_c u_e} C_{u_f s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b r_d} C_{r_c u_e} \Sigma C_{u_e x_g} \Sigma C_{x_g r_c} C_{r_d s_b}) \\
&+ \operatorname{tr}(\Sigma C_{s_b r_d} C_{r_c u_e} C_{u_f x_h} C_{x_g r_c} C_{r_d s_b}) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \\
&+ \operatorname{tr}(\Sigma C_{s_b u_f} C_{u_e x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f s_b}) \operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{r_d u_f} C_{u_e x_g} C_{x_h s_b}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}(\Sigma^2) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} \Sigma C_{r_c u_e} C_{u_f s_b}) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \\
&+ \operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{r_d s_b}) \operatorname{tr}(\Sigma C_{u_e r_c} C_{r_d x_h} C_{x_g u_e})
\end{aligned}
$$

$$+\operatorname{tr}(\Sigma C_{s_b u_f} C_{u_e x_g} \Sigma C_{x_g r_c} C_{r_d s_b})\operatorname{tr}(C_{r_c u_e} C_{u_f r_d})$$

$$+\operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{r_d s_b})\operatorname{tr}(C_{r_c u_e} C_{u_f r_d})\operatorname{tr}(C_{u_e x_g} C_{x_h u_f})$$

$$+\operatorname{tr}(\Sigma C_{s_b u_f} C_{u_e x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f x_h} C_{x_g r_c} C_{r_d s_b})$$

$$+\operatorname{tr}(\Sigma C_{s_b x_h} C_{x_g r_c} C_{r_d u_f} C_{u_e x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f s_b})\Big]$$

$$\mathcal{H}_{12} = \Big[\operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g u_e} \Sigma C_{u_e s_a}) + \operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{u_e x_g} C_{x_h u_f})\operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h u_f} C_{u_e s_a})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a x_g} C_{x_h s_b})\operatorname{tr}(\Sigma C_{u_e s_a} C_{s_b x_h} C_{x_g u_e})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\operatorname{tr}(\Sigma C_{x_g u_e} C_{u_f s_b} C_{s_a x_g})$$

$$+\operatorname{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} \Sigma C_{s_a u_e} \Sigma C_{u_e x_g} C_{x_h r_d}) + \operatorname{tr}(\Sigma C_{r_d u_f} C_{u_e x_g} \Sigma C_{x_g s_a} \Sigma C_{s_a u_e} C_{u_f r_d})$$

$$+\operatorname{tr}(\Sigma C_{r_d s_b} C_{s_a u_e} \Sigma C_{u_e x_g} \Sigma C_{x_g s_a} C_{s_b r_d})$$

$$+\operatorname{tr}(\Sigma C_{r_d s_b} C_{s_a u_e} C_{u_f x_h} C_{x_g s_a} C_{s_b r_d})\operatorname{tr}(C_{u_e x_g} C_{x_h u_f})$$

$$+\operatorname{tr}(\Sigma C_{r_d u_f} C_{u_e x_g} C_{x_h s_b} C_{s_a u_e} C_{u_f r_d})\operatorname{tr}(C_{s_a x_g} C_{x_h s_b})$$

$$+\operatorname{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{s_b u_f} C_{u_e x_g} C_{x_h r_d})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\operatorname{tr}(C_{u_e x_g} C_{x_h u_f})\operatorname{tr}(C_{s_a x_g} C_{x_h s_b})$$

$$+\operatorname{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} \Sigma C_{s_a u_e} C_{u_f r_d})\operatorname{tr}(C_{u_e x_g} C_{x_h u_f})$$

$$+\operatorname{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{s_b r_d})\operatorname{tr}(\Sigma C_{u_e s_a} C_{s_b x_h} C_{x_g u_e})$$

$$+\operatorname{tr}(\Sigma C_{r_d u_f} C_{u_e x_g} \Sigma C_{x_g s_a} C_{s_b r_d})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})$$

$$+\operatorname{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{s_b r_d})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\operatorname{tr}(C_{u_e x_g} C_{x_h u_f})$$

$$+\operatorname{tr}(\Sigma C_{r_d u_f} C_{u_e x_g} C_{x_h s_b} C_{s_a u_e} C_{u_f x_h} C_{x_g s_a} C_{s_b r_d})$$

$$+\operatorname{tr}(\Sigma C_{r_d x_h} C_{x_g s_a} C_{s_b u_f} C_{u_e x_g} C_{x_h s_b} C_{s_a u_e} C_{u_f r_d})\Big]$$

$$\mathcal{H}_{13} = \Big[\operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g r_c} \Sigma C_{r_c s_a}) + \operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})\operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h r_d} C_{r_c s_a})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a x_g} C_{x_h s_b})\operatorname{tr}(\Sigma C_{r_c s_a} C_{s_b x_h} C_{x_g r_c})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})\operatorname{tr}(\Sigma C_{x_g r_c} C_{r_d s_b} C_{s_a x_g})$$

$$+\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} \Sigma C_{s_a r_c} \Sigma C_{r_c x_g} C_{x_h u_f}) + \operatorname{tr}(\Sigma C_{u_f r_d} C_{r_c x_g} \Sigma C_{x_g s_a} \Sigma C_{s_a r_c} C_{r_d u_f})$$

$$+\operatorname{tr}(\Sigma C_{u_f s_b} C_{s_a r_c} \Sigma C_{r_c x_g} \Sigma C_{x_g s_a} C_{s_b u_f})$$

$$+\operatorname{tr}(\Sigma C_{u_f s_b} C_{s_a r_c} C_{r_d x_h} C_{x_g s_a} C_{s_b u_f})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})$$

$$+\operatorname{tr}(\Sigma C_{u_f r_d} C_{r_c x_g} C_{x_h s_b} C_{s_a r_c} C_{r_d u_f})\operatorname{tr}(C_{s_a x_g} C_{x_h s_b})$$

$$+\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} C_{s_b r_d} C_{r_c x_g} C_{x_h u_f})\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})\operatorname{tr}(C_{s_a x_g} C_{x_h s_b})$$

$$+\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} \Sigma C_{s_a r_c} C_{r_d u_f})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})$$

$$+\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} C_{s_b u_f})\operatorname{tr}(\Sigma C_{r_c s_a} C_{s_b x_h} C_{x_g r_c})$$

$$+\operatorname{tr}(\Sigma C_{u_f r_d} C_{r_c x_g} \Sigma C_{x_g s_a} C_{s_b u_f})\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})$$

$$+\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} C_{s_b u_f})\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})\operatorname{tr}(C_{r_c x_g} C_{x_h r_d})$$

$$+\operatorname{tr}(\Sigma C_{u_f r_d} C_{r_c x_g} C_{x_h s_b} C_{s_a r_c} C_{r_d x_h} C_{x_g s_a} C_{s_b u_f})$$

$$+\operatorname{tr}(\Sigma C_{u_f x_h} C_{x_g s_a} C_{s_b r_d} C_{r_c x_g} C_{x_h s_b} C_{s_a r_c} C_{r_d u_f})\Big]$$

$$\mathcal{H}_{14} = \Big[\operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma C_{s_a u_e} \Sigma C_{u_e r_c} \Sigma C_{r_c s_a}) + \operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{r_c u_e} C_{u_f r_d})\operatorname{tr}(\Sigma C_{s_a u_e} C_{u_f r_d} C_{r_c s_a})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})\operatorname{tr}(\Sigma C_{r_c s_a} C_{s_b u_f} C_{u_e r_c})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})\operatorname{tr}(\Sigma C_{u_e r_c} C_{r_d s_b} C_{s_a u_e})$$

$$+\operatorname{tr}(\Sigma C_{* u_f} C_{u_e s_a} \Sigma C_{s_a r_c} \Sigma C_{r_c u_e} C_{u_f *}) + \operatorname{tr}(\Sigma C_{* r_d} C_{r_c u_e} \Sigma C_{u_e s_a} \Sigma C_{s_a r_c} C_{r_d *})$$

$$+\operatorname{tr}(\Sigma C_{* s_b} C_{s_a r_c} \Sigma C_{r_c u_e} \Sigma C_{u_e s_a} C_{s_b *})$$

$$+\operatorname{tr}(\Sigma C_{* s_b} C_{s_a r_c} C_{r_d u_f} C_{u_e s_a} C_{s_b *})\operatorname{tr}(C_{r_c u_e} C_{u_f r_d})$$

$$+\operatorname{tr}(\Sigma C_{* r_d} C_{r_c u_e} C_{u_f s_b} C_{s_a r_c} C_{r_d *})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})$$

$$+\operatorname{tr}(\Sigma C_{* u_f} C_{u_e s_a} C_{s_b r_d} C_{r_c u_e} C_{u_f *})\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})$$

$$+\operatorname{tr}(\Sigma^2)\operatorname{tr}(C_{s_a r_c} C_{r_d s_b})\operatorname{tr}(C_{r_c u_e} C_{u_f r_d})\operatorname{tr}(C_{s_a u_e} C_{u_f s_b})$$

$$
\begin{aligned}
&+ \operatorname{tr}(\Sigma C_{*u_f} C_{u_e s_a} \Sigma C_{s_a r_c} C_{r_d *}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}(\Sigma C_{*u_f} C_{u_e s_a} C_{s_b *}) \operatorname{tr}(\Sigma C_{r_c s_a} C_{s_b u_f} C_{u_e r_c}) \\
&+ \operatorname{tr}(\Sigma C_{*r_d} C_{r_c u_e} \Sigma C_{u_e s_a} C_{s_b *}) \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \\
&+ \operatorname{tr}(\Sigma C_{*u_f} C_{u_e s_a} C_{s_b *}) \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \\
&+ \operatorname{tr}(\Sigma C_{*r_d} C_{r_c u_e} C_{u_f s_b} C_{s_a r_c} C_{r_d u_f} C_{u_e s_a} C_{s_b *}) \\
&+ \operatorname{tr}(\Sigma C_{*u_f} C_{u_e s_a} C_{s_b r_d} C_{r_c u_e} C_{u_f s_b} C_{s_a r_c} C_{r_d *}) \Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_{15} = \Big[&\operatorname{tr}(\Sigma C_{s_a x_x} \Sigma C_{x_g u_e} \Sigma C_{u_e r_c} \Sigma C_{r_c s_a}) + \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h u_f} C_{u_e r_c} \Sigma C_{r_c s_a}) \\
&+ \operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h r_d} C_{r_c s_a}) \operatorname{tr}(\Sigma C_{u_e r_c} C_{r_d x_h} C_{x_g u_e}) \\
&+ \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \operatorname{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g u_e} C_{u_f r_d} C_{r_c s_a}) \\
&+ \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(\Sigma C_{r_c s_a} C_{s_b x_h} C_{x_g u_e} \Sigma C_{u_e r_c}) \\
&+ \operatorname{tr}(\Sigma C_{r_c s_a} C_{s_b u_f} C_{u_e r_c}) \operatorname{tr}(\Sigma C_{x_g u_e} C_{u_f s_b} C_{s_a x_g}) \\
&+ \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \operatorname{tr}(\Sigma C_{u_e r_c} C_{r_d s_b} C_{s_a x_g} \Sigma C_{x_g u_e}) \\
&+ \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \operatorname{tr}(C_{r_c u_e} C_{u_f x_h} C_{x_g s_a} C_{s_b r_d}) \\
&+ \operatorname{tr}(C_{s_a r_c} C_{r_d x_h} C_{x_g s_a} C_{s_b u_f} C_{u_e x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f s_b}) \\
&+ \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{s_a r_c} C_{r_d u_f} C_{u_e x_g} C_{x_h s_b}) \\
&+ \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h r_d} C_{r_c s_a}) \\
&+ \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \operatorname{tr}(\Sigma C_{r_c s_a} C_{s_b u_f} C_{u_e r_c}) \\
&+ \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(\Sigma C_{u_e r_c} C_{r_d x_h} C_{x_g u_e}) \\
&+ \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \operatorname{tr}(\Sigma C_{x_g u_e} C_{u_f s_b} C_{s_a x_g}) \\
&+ \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \operatorname{tr}(C_{r_c u_e} C_{u_f r_d}) \operatorname{tr}(C_{u_e x_g} C_{x_h u_f}) \operatorname{tr}(C_{x_g s_a} C_{s_b x_h}) \\
&+ \operatorname{tr}(C_{s_a r_c} C_{r_d s_b}) \operatorname{tr}(C_{r_c u_e} C_{u_f x_h} C_{x_g s_a} C_{s_b u_f} C_{u_e x_g} C_{x_h r_d}) \\
&+ \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{s_a r_c} C_{r_d u_f} C_{u_e x_g} C_{x_h r_d} C_{r_c u_e} C_{u_f s_b}) \Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_{16} = \Big[& \mathrm{tr}(\Sigma C_{s_a x_x} \Sigma C_{u_e x_g} \Sigma C_{x_g r_c} \Sigma C_{r_c s_a}) + \mathrm{tr}(C_{x_g u_e} C_{u_f x_h}) \mathrm{tr}(\Sigma C_{s_a u_e} C_{u_f x_h} C_{x_g r_c} \Sigma C_{r_c s_a}) \\
& + \mathrm{tr}(\Sigma C_{s_a u_e} C_{u_f r_d} C_{r_c s_a}) \mathrm{tr}(\Sigma C_{x_g r_c} C_{r_d u_f} C_{u_e x_g}) \\
& + \mathrm{tr}(C_{r_c x_g} C_{x_h r_d}) \mathrm{tr}(\Sigma C_{s_a u_e} \Sigma C_{u_e x_g} C_{x_h r_d} C_{r_c s_a}) \\
& + \mathrm{tr}(C_{s_a u_e} C_{u_f s_b}) \mathrm{tr}(\Sigma C_{r_c s_a} C_{s_b u_f} C_{u_e x_g} \Sigma C_{x_g r_c}) \\
& + \mathrm{tr}(\Sigma C_{r_c s_a} C_{s_b x_h} C_{x_g r_c}) \mathrm{tr}(\Sigma C_{u_e x_g} C_{x_h s_b} C_{s_a u_e}) \\
& + \mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \mathrm{tr}(\Sigma C_{x_g r_c} C_{r_d s_b} C_{s_a u_e} \Sigma C_{u_e x_g}) \\
& + \mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \mathrm{tr}(C_{x_g u_e} C_{u_f x_h}) \mathrm{tr}(C_{r_c x_g} C_{x_h u_f} C_{u_e s_a} C_{s_b r_d}) \\
& + \mathrm{tr}(C_{s_a r_c} C_{r_d u_f} C_{u_e s_a} C_{s_b x_h} C_{x_g u_e} C_{u_f r_d} C_{r_c x_g} C_{x_h s_b}) \\
& + \mathrm{tr}(C_{r_c x_g} C_{x_h r_d}) \mathrm{tr}(C_{s_a u_e} C_{u_f s_b}) \mathrm{tr}(C_{s_a r_c} C_{r_d x_h} C_{x_g u_e} C_{u_f s_b}) \\
& + \mathrm{tr}(C_{r_c x_g} C_{x_h r_d}) \mathrm{tr}(C_{x_g u_e} C_{u_f x_h}) \mathrm{tr}(\Sigma C_{s_a u_e} C_{u_f r_d} C_{r_c s_a}) \\
& + \mathrm{tr}(C_{s_a u_e} C_{u_f s_b}) \mathrm{tr}(C_{x_g u_e} C_{u_f x_h}) \mathrm{tr}(\Sigma C_{r_c s_a} C_{s_b x_h} C_{x_g r_c}) \\
& + \mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \mathrm{tr}(C_{s_a u_e} C_{u_f s_b}) \mathrm{tr}(\Sigma C_{x_g r_c} C_{r_d u_f} C_{u_e x_g}) \\
& + \mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \mathrm{tr}(C_{r_c x_g} C_{x_h r_d}) \mathrm{tr}(\Sigma C_{u_e x_g} C_{x_h s_b} C_{s_a u_e}) \\
& + \mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \mathrm{tr}(C_{r_c x_g} C_{x_h r_d}) \mathrm{tr}(C_{x_g u_e} C_{u_f x_h}) \mathrm{tr}(C_{u_e s_a} C_{s_b u_f}) \\
& + \mathrm{tr}(C_{s_a r_c} C_{r_d s_b}) \mathrm{tr}(C_{r_c x_g} C_{x_h u_f} C_{u_e s_a} C_{s_b x_h} C_{x_g u_e} C_{u_f r_d}) \\
& + \mathrm{tr}(C_{s_a u_e} C_{u_f s_b}) \mathrm{tr}(C_{s_a r_c} C_{r_d x_h} C_{x_g u_e} C_{u_f r_d} C_{r_c x_g} C_{x_h s_b}) \Big]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}_{17} = \Big[& \mathrm{tr}(\Sigma C_{s_a x_x} \Sigma C_{x_g r_c} \Sigma C_{r_c u_e} \Sigma C_{u_e s_a}) + \mathrm{tr}(C_{r_c x_g} C_{x_h r_d}) \mathrm{tr}(\Sigma C_{s_a x_g} C_{x_h r_d} C_{r_c u_e} \Sigma C_{u_e s_a}) \\
& + \mathrm{tr}(\Sigma C_{s_a x_g} C_{x_h u_f} C_{u_e s_a}) \mathrm{tr}(\Sigma C_{r_c u_e} C_{u_f x_h} C_{x_g r_c}) \\
& + \mathrm{tr}(C_{u_e r_c} C_{r_d u_f}) \mathrm{tr}(\Sigma C_{s_a x_g} \Sigma C_{x_g r_c} C_{r_d u_f} C_{u_e s_a}) \\
& + \mathrm{tr}(C_{s_a x_g} C_{x_h s_b}) \mathrm{tr}(\Sigma C_{u_e s_a} C_{s_b x_h} C_{x_g r_c} \Sigma C_{r_c u_e}) \\
& + \mathrm{tr}(\Sigma C_{u_e s_a} C_{s_b r_d} C_{r_c u_e}) \mathrm{tr}(\Sigma C_{x_g r_c} C_{r_d s_b} C_{s_a x_g})
\end{aligned}
$$

$$+ \operatorname{tr}(C_{s_a u_e} C_{u_f s_b}) \operatorname{tr}(\Sigma C_{r_c u_e} C_{u_f s_b} C_{s_a x_g} \Sigma C_{x_g r_c})$$

$$+ \operatorname{tr}(C_{s_a u_e} C_{u_f s_b}) \operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \operatorname{tr}(C_{u_e r_c} C_{r_d x_h} C_{x_g s_a} C_{s_b u_f})$$

$$+ \operatorname{tr}(C_{s_a u_e} C_{u_f x_h} C_{x_g s_a} C_{s_b r_d} C_{r_c x_g} C_{x_h u_f} C_{u_e r_c} C_{r_d s_b})$$

$$+ \operatorname{tr}(C_{u_e r_c} C_{r_d u_f}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{s_a u_e} C_{u_f r_d} C_{r_c x_g} C_{x_h s_b})$$

$$+ \operatorname{tr}(C_{u_e r_c} C_{r_d u_f}) \operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \operatorname{tr}(\Sigma C_{s_a x_g} C_{x_h u_f} C_{u_e s_a})$$

$$+ \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \operatorname{tr}(\Sigma C_{u_e s_a} C_{s_b r_d} C_{r_c u_e})$$

$$+ \operatorname{tr}(C_{s_a u_e} C_{u_f s_b}) \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(\Sigma C_{r_c u_e} C_{u_f x_h} C_{x_g r_c})$$

$$+ \operatorname{tr}(C_{s_a u_e} C_{u_f s_b}) \operatorname{tr}(C_{u_e r_c} C_{r_d u_f}) \operatorname{tr}(\Sigma C_{x_g r_c} C_{r_d s_b} C_{s_a x_g})$$

$$+ \operatorname{tr}(C_{s_a u_e} C_{u_f s_b}) \operatorname{tr}(C_{u_e r_c} C_{r_d u_f}) \operatorname{tr}(C_{r_c x_g} C_{x_h r_d}) \operatorname{tr}(C_{x_g s_a} C_{s_b x_h})$$

$$+ \operatorname{tr}(C_{s_a u_e} C_{u_f s_b}) \operatorname{tr}(C_{u_e r_c} C_{r_d x_h} C_{x_g s_a} C_{s_b r_d} C_{r_c x_g} C_{x_h u_f})$$

$$+ \operatorname{tr}(C_{s_a x_g} C_{x_h s_b}) \operatorname{tr}(C_{s_a u_e} C_{u_f r_d} C_{r_c x_g} C_{x_h u_f} C_{u_e r_c} C_{r_d s_b})\Big].$$

Proof. By definition of $V_{s_a s_b}(i,j)$,

$$
\begin{aligned}
&E\Big\{V_{s_a s_b}(i,j)V_{r_c r_d}(i,j)V_{u_e u_f}(i,j)V_{x_g x_h}(i,j)\Big\}\\
&\quad = E\Big\{(Y_{is_a}^T Y_{js_b})^2(Y_{ir_c}^T Y_{jr_d})^2(Y_{iu_e}^T Y_{ju_f})^2(Y_{ix_g}^T Y_{jx_h})^2\Big\}\\
&\quad = E\Big\{(\sum_k Y_{is_a k}Y_{js_b k})^2(\sum_m Y_{ir_c m}Y_{jr_d m})^2(\sum_o Y_{iu_e o}Y_{ju_f o})^2(\sum_q Y_{ix_g q}Y_{jx_h q})^2\Big\}\\
&\quad = \sum_{\mathcal{C}} E\Big\{Y_{is_a k}Y_{js_b k}Y_{is_a l}Y_{js_b l}Y_{ir_c m}Y_{jr_d m}Y_{ir_c n}Y_{jr_d n}Y_{iu_e o}Y_{ju_f o}Y_{iu_e p}Y_{ju_f p}\\
&\qquad \times Y_{ix_g q}Y_{jx_h q}Y_{ix_g w}Y_{jx_h w}\Big\}\\
&\quad = \sum_{\mathcal{C}} E\Big\{Y_{is_a k}Y_{is_a l}Y_{ir_c m}Y_{ir_c n}Y_{iu_e o}Y_{iu_e p}Y_{ix_g q}Y_{ix_g w}\Big\}\\
&\qquad \times E\Big\{Y_{js_b k}Y_{js_b l}Y_{jr_d m}Y_{jr_d n}Y_{ju_f o}Y_{ju_f p}Y_{jx_h q}Y_{jx_h w}\Big\}, \qquad (3.18)
\end{aligned}
$$

where $\mathcal{C}$ represents the summation over the p components of the vector $Y_{..}$ for k, l, m, n, o, p, q, and w. For each of the expectation terms in (3.18) we apply lemma 7 and sum over the set $\mathcal{C}$. After some tedious algebra it follows that

$$
\begin{aligned}
E\Big\{V_{s_a s_b}(i,j)V_{r_c r_d}(i,j)V_{u_e u_f}(i,j)V_{x_g x_h}(i,j)\Big\} &= \mathcal{H}_1+\mathcal{H}_2+\mathcal{H}_3+\mathcal{H}_4+\mathcal{H}_5+\mathcal{H}_6\\
&\quad +\mathcal{H}_7+\mathcal{H}_8+\mathcal{H}_9+\mathcal{H}_{10}+\mathcal{H}_{11}+\mathcal{H}_{12}\\
&\quad +\mathcal{H}_{13}+\mathcal{H}_{14}+\mathcal{H}_{15}+\mathcal{H}_{16}+\mathcal{H}_{17}.
\end{aligned}
$$

□

3.8.2 Proofs of theorems

In this section we provide proofs for the theorems given in Chapter 3. Without loss of generality, assume $\mu_t = 0$ for all $t \in \{1, \ldots, T\}$, since the test statistic, $\hat{D}_{nt}$, is invariant with respect to μ_t.

Proof of Theorem 7. With the addition of Condition 4 as an assumption, the proof is similar to the proof of Theorem 2. Condition (a) and Condition (b) hold in the proof of Theorem 2, and the martingale central limit theorem holds as all required terms are smaller order with T diverging. $\square$

Proof of Theorem 8. To establish the asymptotic distribution of $\mathcal{M}_n$ under H_0, we must show convergence of the finite-dimensional distributions and the tightness of the stochastic process $\max_{t\in\mathcal{T}} \sigma_{nt,0}^{-1}\hat{D}_{nt}$. The joint asymptotic normality of $(\sigma_{nt_1,0}^{-1}\hat{D}_{nt_1}, \ldots, \sigma_{nt_c,0}^{-1}\hat{D}_{nt_c})^{\mathrm{T}}$ for $t_1 < \cdots < t_c$ is nearly identical to the proof in 2.7 when T is considered finite. Thus, it remains for us to show the tightness of $\max_{t\in\mathcal{T}} \sigma_{nt,0}^{-1}\hat{D}_{nt}$ so as to conclude $\mathcal{M}_n$ converges to $\max_{t\in\mathcal{T}} Z_t$, where Z_t is a Gaussian process with mean 0 and correlation R_z

By definition, $\hat{D}_{nt} = \hat{D}_{nt,0} + \hat{D}_{nt,2} - 2\hat{D}_{nt,1}$, where

$$\hat{D}_{nt,k} = \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\Big(U_{s_1s_1,k} + U_{s_2s_2,k} - U_{s_1s_2,k} - U_{s_2s_1,k}\Big)$$

for $k \in \{0,1,2\}$. Furthermore, by Lemmas 3 and 4 in 2.7.1, $\hat{D}_{nt,1} = o_p(\hat{D}_{nt,0})$ and $\hat{D}_{nt,2} = o_p(\hat{D}_{nt,0})$. Therefore, to show the tightness of $\max_{1\le t<T} \sigma_{nt,0}^{-1}\hat{D}_{nt}$ we can focus on the term $\hat{D}_{nt,0}$ where

$$\hat{D}_{nt,0} = \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\Big(U_{s_1s_10} + U_{s_2s_20} - U_{s_1s_20} - U_{s_2s_10}\Big). \tag{3.19}$$

Let $t = [T\nu]$, where $\nu = j/T$ $(j = 1, \ldots, T-1)$. Define $G_n(\nu)$ as follows:

$$G_n(\nu) = \frac{\sqrt{n(n-1)}}{\mathrm{tr}(\Sigma^2)T^{\frac{3}{2}}}\hat{D}_{n[T\nu],0} \tag{3.20}$$

where the term preceding $\hat{D}_{n[T\nu],0}$ is the order of $\sigma_{nt,0}^{-1}$ in terms of n, p, and T. Thus, to show the tightness of $\max_{t\in\mathcal{T}} \sigma_{nt,0}^{-1}\hat{D}_{nt}$ it is equivalent to show the tightness of (3.20).

Consider the difference $G_n(\eta) - G_n(\nu)$, such that $\eta > \nu$. Let $\eta = i/T$ $(i = 1, \ldots, T-1)$. By definition, $U_{s_1 s_1 0} = \{n(n-1)\}^{-1} \sum_{i \neq j} (Y_{is_1}^T Y_{js_1})^2$. Thus, for $\eta, \nu \in (0,1)$,

$$G_n(\eta) - G_n(\nu) = \frac{\sqrt{n(n-1)}}{\operatorname{tr}(\Sigma^2) T^{\frac{3}{2}}} \Big(\sum_{s_1=1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} W_{s_1 s_2} - \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{T} W_{s_1 s_2} \Big),$$

$$= \frac{\sqrt{n(n-1)}}{\operatorname{tr}(\Sigma^2) T^{\frac{3}{2}}} \Big(\sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} W_{s_1 s_2} - \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{[T\eta]} W_{s_1 s_2} \Big),$$

$$= \frac{\sqrt{n(n-1)}}{\operatorname{tr}(\Sigma^2) T^{\frac{3}{2}}} \Big[\sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} \Big\{ \frac{1}{n(n-1)} \sum_{i \neq j}^{n} \tilde{W}_{s_1 s_2} \Big\} - \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{[T\eta]} \Big\{ \frac{1}{n(n-1)} \sum_{i \neq j}^{n} \tilde{W}_{s_1 s_2} \Big\} \Big],$$

$$= \frac{1}{\sqrt{n(n-1)} \operatorname{tr}(\Sigma^2) T^{\frac{3}{2}}} \sum_{i \neq j}^{n} \Big(\sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} \tilde{W}_{s_1 s_2} - \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{[T\eta]} \tilde{W}_{s_1 s_2} \Big),$$

$$= \frac{1}{\sqrt{n(n-1)} \operatorname{tr}(\Sigma^2) T^{\frac{3}{2}}} \sum_{i \neq j}^{n} f(i,j), \tag{3.21}$$

where $\tilde{W}_{s_1 s_2} = (Y_{is_1}^T Y_{js_1})^2 + (Y_{is_2}^T Y_{js_2})^2 - (Y_{is_1}^T Y_{js_2})^2 - (Y_{is_2}^T Y_{js_1})^2$, and $f(i,j) = \sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} \tilde{W}_{s_1 s_2} - \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{[T\eta]} \tilde{W}_{s_1 s_2}$. We will bound the fourth moment of (3.21) to ultimately show the tightness of (3.20). First, we compute some moments of f for various indices.

Under the null hypothesis and for $i \neq j$,

$$E\Big\{f(i,j)\Big\} = E\Big(\sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} \tilde{W}_{s_1 s_2} - \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{[T\eta]} \tilde{W}_{s_1 s_2}\Big),$$

$$= \sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} \Big[\mathrm{tr}(\Sigma_{s_1}^2) + \mathrm{tr}(\Sigma_{s_2}^2) - 2tr(\Sigma_{s_1}\Sigma_{s_2})\Big] - \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{[T\eta]} \Big[\mathrm{tr}(\Sigma_{s_1}^2) + \mathrm{tr}(\Sigma_{s_2}^2) - 2tr(\Sigma_{s_1}\Sigma_{s_2})\Big],$$

$$= 0, \tag{3.22}$$

since $\mathrm{tr}(\Sigma_{s_1}^2) + \mathrm{tr}(\Sigma_{s_2}^2) - 2tr(\Sigma_{s_1}\Sigma_{s_2}) = \mathrm{tr}(\Sigma^2) + \mathrm{tr}(\Sigma^2) - 2tr(\Sigma^2) = 0$. Define the following notation for the double summation: $\sum_{\mathcal{S}^1} \equiv \sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T}$; $\sum_{\mathcal{R}^1} \equiv \sum_{r_1=[T\nu]+1}^{[T\eta]} \sum_{r_2=[T\eta]+1}^{T}$; $\sum_{\mathcal{S}^2} \equiv \sum_{s_1=1}^{[T\nu]} \sum_{s_2=[T\nu]+1}^{[T\eta]}$; $\sum_{\mathcal{R}^2} \equiv \sum_{r_1=1}^{[T\nu]} \sum_{r_2=[T\nu]+1}^{[T\eta]}$. The second moment under the null hypothesis is given by

$$E\Big\{f(i,j)f(i,j)\Big\} = E\Big\{\Big(\sum_{\mathcal{S}^1} \tilde{W}_{s_1 s_2} - \sum_{\mathcal{S}^2} \tilde{W}_{s_1 s_2}\Big)\Big(\sum_{\mathcal{R}^1} \tilde{W}_{s_1 s_2} - \sum_{\mathcal{R}^2} \tilde{W}_{s_1 s_2}\Big)\Big\},$$

$$= \sum_{x,y=1}^{2} (-1)^{|x-y|} \sum_{\mathcal{S}^x} \sum_{\mathcal{R}^y} E\Big[\tilde{W}_{s_1 s_2} \tilde{W}_{r_1 r_2}\Big],$$

$$= \sum_{x,y=1}^{2} (-1)^{|x-y|} \sum_{\mathcal{S}^x} \sum_{\mathcal{R}^y} \sum_{a,b,c,d=1}^{2} (-1)^{|a-b|+|c-d|} E\Big[V_{s_a s_b}(i,j) V_{r_c r_d}(i,j)\Big],$$

where $V_{s_a s_b}(i,j) = (Y_{is_a}^T Y_{js_b})^2$. Under the null hypothesis,

$$E\Big[V_{s_a s_b}(i,j) V_{r_c r_d}(i,j)\Big] = 2\mathrm{tr}^2(C_{s_b r_d} C_{r_c s_a}) + 2tr(C_{r_c s_a} C_{s_b r_d} C_{r_c s_a} C_{s_b r_d}) + 2\mathrm{tr}(\Sigma C_{r_d s_b} \Sigma C_{s_b r_d}) + 2\mathrm{tr}(\Sigma C_{r_c s_a} \Sigma C_{s_a r_c}) + \mathrm{tr}^2(\Sigma^2).$$

Therefore, under Condition 1,

$$E\Big\{f(i,j)f(i,j)\Big\} = C\sum_{x,y=1}^{2}(-1)^{|x-y|}\sum_{\mathcal{S}^x}\sum_{\mathcal{R}^y}\sum_{a,b,c,d=1}^{2}(-1)^{|a-b|+|c-d|}\mathrm{tr}^2(C_{s_b r_d}C_{r_c s_a}) \tag{3.23}$$

for some constant C.

Next, consider mutually different indices i, j, k. Thus,

$$E\Big\{f(i,j)f(i,k)\Big\} = \sum_{x,y=1}^{2}(-1)^{|x-y|}\sum_{\mathcal{S}^x}\sum_{\mathcal{R}^y}\sum_{a,b,c,d=1}^{2}(-1)^{|a-b|+|c-d|}E\Big[V_{s_a s_b}(i,j)V_{r_c r_d}(i,k)\Big]. \tag{3.24}$$

Under the null hypothesis,

$$E\Big[V_{s_a s_b}(i,j)V_{r_c r_d}(i,k)\Big] = \mathrm{tr}^2(\Sigma^2) + 2\mathrm{tr}(\Sigma C_{r_c s_a}\Sigma C_{s_a r_c}).$$

Hence,

$$\sum_{a,b,c,d=1}^{2}(-1)^{|a-b|+|c-d|}\Big\{\mathrm{tr}^2(\Sigma^2) + 2\mathrm{tr}(\Sigma C_{r_c s_a}\Sigma C_{s_a r_c})\Big\} = 0. \tag{3.25}$$

Therefore, $E\Big\{f(i,j)f(i,k)\Big\} = 0$. Lastly, if we consider the mutually different indices i, j, k, l, then $E\Big\{f(i,j)f(k,l)\Big\} = 0$ due to independence and the fact that $E\Big\{f(i,j)\Big\} = 0$.

Consider the difference $G_n(\eta) - G_n(\nu)$ squared.

$$\{G_n(\eta) - G_n(\nu)\}^2 = \{n(n-1)\mathrm{tr}^2(\Sigma^2)T^3\}^{-1}\Big\{\sum_{i\neq j} f(i,j)\Big\}^2,$$

$$\begin{aligned} &= 2\{n(n-1)\mathrm{tr}^2(\Sigma^2)T^3\}^{-1}\sum_{i\neq j} f(i,j)f(i,j) \\ &\quad + 4\{n(n-1)\mathrm{tr}^2(\Sigma^2)T^3\}^{-1}\sum_{i\neq j\neq k} f(i,j)f(i,k) \\ &\quad + \{n(n-1)\mathrm{tr}^2(\Sigma^2)T^3\}^{-1}\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l). \end{aligned}$$

For any real numbers a, b, and c, $(a+b+c)^2 \leq 4a^2+4b^2+2c^2$. Thus,

$$\begin{aligned}\{G_n(\eta)-G_n(\nu)\}^4 \leq\ & 16\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}\Big\{\sum_{i\neq j} f(i,j)f(i,j)\Big\}^2 \\ & +64\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}\Big\{\sum_{i\neq j\neq k} f(i,j)f(i,k)\Big\}^2 \\ & +2\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)\Big\}^2.\end{aligned}$$

Taking the expectation of both sides of the above inequality it follows that

$$\begin{aligned}E\Big[\{G_n(\eta)-G_n(\nu)\}^4\Big] \leq\ & 16\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}E\Big[\Big\{\sum_{i\neq j} f(i,j)f(i,j)\Big\}^2\Big] \\ & +64\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}E\Big[\Big\{\sum_{i\neq j\neq k} f(i,j)f(i,k)\Big\}^2\Big] \\ & +2\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}E\Big[\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)\Big\}^2\Big], \\ \equiv\ & I_1+I_2+I_3. \end{aligned} \tag{3.26}$$

To bound the expectation in (3.26) we need the order of I_1, I_2, and I_3. Thus, we need to expand multiple summations where the summation is across multiple non-identical indices. First, consider the possible indices for expanding the term inside the expectation for I_1 in (3.26). Consider

$$\Big\{\sum_{i\neq j} f(i,j)f(i,j)\Big\}^2 = \sum_{i\neq j}\sum_{i_1\neq j_1} f(i,j)f(i,j)f(i_1,j_1)f(i_1,j_1). \tag{3.27}$$

Let $D_c=\{i,j\}\cup\{i_1,j_1\}$ be the set of indices that are not equivalent to each other where c represents the number of indices that are equivalent to each other in two sets $\{i,j\}$ and $\{i_1,j_1\}$. If there are no equivalent indices, then

$$D_0=\{(i,j,i_1,j_1)\}.$$

Hence, the summation over D_0 is given by

$$\sum_{i \neq i_1 \neq j \neq j_1} f(i,j)f(i,j)f(i_1,j_1)f(i_1,j_1) \tag{3.28}$$

If there is one equivalent index, then

$$D_1 = \{(i = i_1, j, j_1), (i = j_1, j, i_1), (i, j = i_1, j_1), (i, j = j_1, i_1)\}.$$

Let $\tilde{D_1}$ be the set with one equivalent index that produces a unique combination of $f(i,j)f(i,j)f(i_1,j_1)f(i_1,j_1)$.

$$\tilde{D_1} = \{(i = i_1, j, j_1)\}.$$

Hence, the summation over D_1 is equivalent to

$$\sum_{i \neq j \neq j_1} 4f(i,j)f(i,j)f(i,j_1)f(i,j_1) \tag{3.29}$$

If there are two equivalent indices, then

$$D_2 = \{(i = i_1, j = j_1), (i = j_1, j = i_1)\}.$$

Hence, the summation over D_2 is given by

$$\sum_{i \neq j} 2f(i,j)f(i,j)f(i,j)f(i,j). \tag{3.30}$$

As a result, from (3.28) – (3.30),

$$\begin{aligned} E\left[\left\{\sum_{i \neq j} f(i,j)f(i,j)\right\}^2\right] &= E\left\{\sum_{i \neq i_1 \neq j \neq j_1} f(i,j)f(i,j)f(i_1,j_1)f(i_1,j_1)\right\} \\ &\quad + 4E\left\{\sum_{i \neq j \neq j_1} f(i,j)f(i,j)f(i,j_1)f(i,j_1)\right\} \\ &\quad + 2E\left\{\sum_{i \neq j} f(i,j)f(i,j)f(i,j)f(i,j)\right\}. \end{aligned} \tag{3.31}$$

Thus,

$$I_1 = 16\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}E\Bigg[\Big\{\sum_{i\neq j} f(i,j)f(i,j)\Big\}^2\Bigg]$$

$$= 16\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}E\Big\{\sum_{i\neq i_1\neq j\neq j_1} f(i,j)f(i,j)f(i_1,j_1)f(i_1,j_1)\Big\} \tag{3.32}$$

$$+ 64\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}E\Big\{\sum_{i\neq j\neq j_1} f(i,j)f(i,j)f(i,j_1)f(i,j_1)\Big\}$$

$$+ 32\{n^2(n-1)^2tr^4(\Sigma^2)T^6\}^{-1}E\Big\{\sum_{i\neq j} f(i,j)f(i,j)f(i,j)f(i,j)\Big\}$$

$$\equiv R_1 + R_2 + R_3. \tag{3.33}$$

We now show the order for each of R_1, R_2, and R_3 in terms of n, p, and T. For R_1, consider the order for $E\{f(i,j)f(i,j)f(i_1,j_1)f(i_1,j_1)\}$ for the mutually different indices.

$$E\Big\{f(i,j)f(i,j)f(i_1,j_1)f(i_1,j_1)\Big\} = E\Big\{f(i,j)f(i,j)\Big\}E\Big\{f(i_1,j_1)f(i_1,j_1)\Big\}$$

$$= \Bigg[C\sum_{x,y=1}^{2}(-1)^{|x-y|}\sum_{\mathcal{S}^x}\sum_{\mathcal{R}^y} \times \sum_{a,b,c,d=1}^{2}(-1)^{|a-b|+|c-d|}\mathrm{tr}^2(C_{s_b r_d}C_{r_c s_a})\Bigg]^2$$

$$\asymp C\Bigg[\mathrm{tr}^2(\Sigma^2)\Big\{T^2([T\eta]-[T\nu])\Big\}\Bigg]^2. \tag{3.34}$$

for some constant C. Therefore,

$$R_1 \asymp \frac{C[n^2(n-1)^2]}{n^2(n-1)^2tr^4(\Sigma^2)T^6}\Bigg[\mathrm{tr}^2(\Sigma^2)\Big\{T^2([T\eta]-[T\nu])\Big\}\Bigg]^2,$$

$$\asymp C\frac{([T\eta]-[T\nu])^2}{T^2}, \tag{3.35}$$

for some constant C. Next, consider the term R_3. From lemma 7, $E\{V_{s_a s_b}(i,j)V_{r_c r_d}(i,j)V_{u_e u_f}(i,j)V_{x_g x_h}(i,j)\}$ was calculated up to a constant. Thus, $E\Big\{f(i,j)f(i,j)f(i,j)f(i,j)\Big\}$ is given by

$$\sum_{w,x,y,z=1}^{2}(-1)^{|w-x|+|y-z|}\sum_{\mathcal{S}^w}\sum_{\mathcal{R}^x}\sum_{\mathcal{U}^y}\sum_{\mathcal{X}^z}$$
$$\times\sum_{a,b,c,d,e,f,g,h=1}^{2}(-1)^{|a-b|+|c-d|+|e-f|+|g-h|}E\Big\{V_{s_a s_b}(i,j)V_{r_c r_d}(i,j)$$
$$\times V_{u_e u_f}(i,j)V_{x_g x_h}(i,j)\Big\}.$$

Under the null hypothesis,

$$E\Big\{f(i,j)f(i,j)f(i,j)f(i,j)\Big\}\asymp C\Big[\mathrm{tr}^2(\Sigma^2)\Big\{T^2([T\eta]-[T\nu])\Big\}\Big]^2.$$

Therefore,

$$R_3\asymp\frac{C}{[n(n-1)]\mathrm{tr}^4(\Sigma^2)T^6}\Big[\mathrm{tr}^2(\Sigma^2)\Big\{T^2([T\eta]-[T\nu])\Big\}\Big]^2,$$

and thus $R_3=o(R_1)$. For the final term in I_1, R_2, consider the order of $E\{f(i,j)f(i,j)f(i,j_1)f(i,j_1)\}$. By the Cauchy-Schwarz inequality

$$E\Big\{f(i,j)f(i,j)f(i,j_1)f(i,j_1)\Big\}\le\Big[E\Big\{f(i,j)f(i,j)f(i,j)f(i,j)\Big\}\Big]^{1/2}$$
$$\times\Big[E\Big\{f(i,j_1)f(i,j_1)f(i,j_1)f(i,j_1)\Big\}\Big]^{1/2}$$

$$\asymp\mathcal{O}\Big(E\Big\{f(i,j)f(i,j)f(i,j)f(i,j)\Big\}\Big).\tag{3.36}$$

Therefore, based on the above results for R_3 it follows that $R_2=o(R_1)$. As a result,

$$I_1\le C\frac{([T\eta]-[T\nu])^2}{T^2},\tag{3.37}$$

for some constant C.

Next, we investigate the order of I_2. Consider the possible indices for expanding

$$\Big\{\sum_{i\neq j\neq k} f(i,j)f(i,k)\Big\}^2 = \sum_{i\neq j\neq k}\sum_{i_1\neq j_1\neq k_1} f(i,j)f(i,k)f(i_1,j_1)f(i_1,k_1). \tag{3.38}$$

Let $E_c = \{i,j,k\}\cup\{i_1,j_1,k_1\}$ be the set of indices that are not equivalent to each other, where c represents the number of indices that are equivalent to each other in two sets $\{i,j,k\}$ and $\{i_1,j_1,k_1\}$. If there are no equivalent indices, then

$$E_0 = \{(i,j,k,i_1,j_1,k_1)\}.$$

Hence, the summation over E_0 is given by

$$\sum_{i\neq i_1\neq j\neq j_1\neq k\neq k_1} f(i,j)f(i,k)f(i_1,j_1)f(i_1,k_1). \tag{3.39}$$

If there is one equivalent index, then

$$\begin{aligned} E_1 = \{&(i=i_1,j,k,j_1,k_1),(i=j_1,j,k,i_1,k_1),(i=k_1,j,k,i_1,j_1),\\ &(i,j=i_1,k,j_1,k_1),(i,j=j_1,k,i_1,k_1),(i,j=k_1,k,i_1,j_1),\\ &(i,j,k=i_1,j_1,k_1),(i,j,k=j_1,i_1,k_1),(i,j,k=k_1,i_1,j_1)\}.\end{aligned}$$

Let $\tilde{E_1}$ be the set with one equivalent index that produces a unique combination of $f(i,j)f(i,k)f(i_1,j_1)f(i_1,k_1)$, such that

$$\tilde{E_1} = \{(i=i_1,j,k,j_1,k_1),(i=j_1,j,k,i_1,k_1),(i,j=j_1,k,i_1,k_1)\}.$$

Hence, the summation over E_1 is equivalent to

$$\begin{aligned} &\sum_{i\neq j\neq j_1\neq k\neq k_1} f(i,j)f(i,k)f(i,j_1)f(i,k_1) \\ &\quad+\sum_{i\neq i_1\neq j\neq k\neq k_1} 4f(i,j)f(i,k)f(i_1,i)f(i_1,k_1)\\ &\quad+\sum_{i\neq i_1\neq j\neq k\neq k_1} 4f(i,j)f(i,k)f(i_1,j)f(i_1,k_1)\end{aligned} \tag{3.40}$$

If there are two equivalent indices, then

$$\begin{aligned}E_2 = \{&(i=i_1, j=j_1, k, k_1), (i=i_1, j=k_1, k, j_1), (i=j_1, j=i_1, k, k_1),\\ &(i=j_1, j=k_1, k, i_1), (i=k_1, j=i_1, k, j_1), (i=k_1, j=j_1, k, i_1),\\ &(i=i_1, k=j_1, j, k_1), (i=i_1, k=k_1, j, j_1), (i=j_1, k=i_1, j, k_1),\\ &(i=j_1, k=k_1, k, i_1), (i=k_1, k=i_1, j, j_1), (i=k_1, k=j_1, j, i_1),\\ &(j=i_1, k=j_1, i, k_1), (j=i_1, k=k_1, i, j_1), (j=j_1, k=i_1, i, k_1),\\ &(j=j_1, k=k_1, i, i_1), (j=k_1, k=i_1, i, j_1), (j=k_1, k=j_1, i, i_1)\}.\end{aligned}$$

Let $\tilde{E_2}$ be the set with two equivalent indices that produces a unique combination of $f(i,j)f(i,k)f(i_1,j_1)f(i_1,k_1)$.

$\tilde{E_2} = \{(i=i_1, j=j_1, k, k_1), (i=j_1, j=i_1, k, k_1), (i=j_1, j=k_1, k, i_1), (j=j_1, k=k_1, i, i_1)\}$.

Hence, the summation over E_2 is equivalent to

$$\begin{aligned}&\sum_{i\neq j\neq k\neq k_1} 4f(i,j)f(i,k)f(i,j)f(i,k_1)\\ &+ \sum_{i\neq j\neq k\neq k_1} 4f(i,j)f(i,k)f(j,i)f(j,k_1)\\ &+ \sum_{i\neq i_1\neq j\neq k} 8f(i,j)f(i,k)f(i_1,i)f(i_1,j)\\ &+ \sum_{i\neq i_1\neq j\neq k} 2f(i,j)f(i,k)f(i_1,j)f(i_1,k).\end{aligned} \tag{3.41}$$

Lastly, if there are three equivalent indices, then

$$\begin{aligned}E_3 = \{&(i=i_1, j=j_1, k=k_1), (i=i_1, j=k_1, k=j_1), (i=j_1, j=i_1, k=k_1),\\ &(i=j_1, j=k_1, k=i_1), (i=k_1, j=j_1, k=i_1), (i=k_1, j=i_1, k=j_1)\}.\end{aligned}$$

Let $\tilde{E_3}$ be the set with two equivalent indices that produces a unique combination of $f(i,j)f(i,k)f(i_1,j_1)f(i_1,k_1)$ such that

$$\tilde{E_3} = \{(i=i_1, j=j_1, k=k_1), (i=j_1, j=i_1, k=k_1)\}.$$

Hence, the summation over E_3 is equivalent to

$$\begin{aligned}&\sum_{i\neq j\neq k} 2f(i,j)f(i,k)f(i,j)f(i,k) \\ &\quad+ \sum_{i\neq j\neq k} 4f(i,j)f(i,k)f(j,i)f(j,k)\end{aligned} \tag{3.42}$$

As a result, from (3.39) – (3.42),

$$\begin{aligned}E\Bigg[\Bigg\{\sum_{i\neq j\neq k} f(i,j)f(i,k)\Bigg\}^2\Bigg] &= E\Bigg\{\sum_{i\neq i_1\neq j\neq j_1\neq k\neq k_1} f(i,j)f(i,k)f(i_1,j_1)f(i_1,k_1)\Bigg\} \\ &+ E\Bigg\{\sum_{i\neq j\neq j_1\neq k\neq k_1} f(i,j)f(i,k)f(i,j_1)f(i,k_1)\Bigg\} \\ &+ 4E\Bigg\{\sum_{i\neq i_1\neq j\neq k\neq k_1} f(i,j)f(i,k)f(i_1,i)f(i_1,k_1)\Bigg\} \\ &+ 4E\Bigg\{\sum_{i\neq i_1\neq j\neq k\neq k_1} f(i,j)f(i,k)f(i_1,j)f(i_1,k_1)\Bigg\} \\ &+ 4E\Bigg\{\sum_{i\neq j\neq k\neq k_1} f(i,j)f(i,k)f(i,j)f(i,k_1)\Bigg\} \\ &+ 4E\Bigg\{\sum_{i\neq j\neq k\neq k_1} f(i,j)f(i,k)f(j,i)f(j,k_1)\Bigg\} \\ &+ 8E\Bigg\{\sum_{i\neq i_1\neq j\neq k} f(i,j)f(i,k)f(i_1,i)f(i_1,j)\Bigg\} \\ &+ 2E\Bigg\{\sum_{i\neq i_1\neq j\neq k} f(i,j)f(i,k)f(i_1,j)f(i_1,k)\Bigg\} \\ &+ 2E\Bigg\{\sum_{i\neq j\neq k} f(i,j)f(i,k)f(i,j)f(i,k)\Bigg\} \\ &+ 4E\Bigg\{\sum_{i\neq j\neq k} f(i,j)f(i,k)f(j,i)f(j,k)\Bigg\}.\end{aligned} \tag{3.43}$$

Thus,

$$I_2 = 64\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\Bigg[\bigg\{\sum_{i\neq j\neq k} f(i,j)f(i,k)\bigg\}^2\Bigg],$$

$$\begin{aligned}
&= 64\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq i_1\neq j\neq j_1\neq k\neq k_1} f(i,j)f(i,k)f(i_1,j_1)f(i_1,k_1)\bigg\}\\
&\quad + 64\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq j\neq j_1\neq k\neq k_1} f(i,j)f(i,k)f(i,j_1)f(i,k_1)\bigg\}\\
&\quad + 256\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq i_1\neq j\neq k\neq k_1} f(i,j)f(i,k)f(i_1,i)f(i_1,k_1)\bigg\}\\
&\quad + 256\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq i_1\neq j\neq k\neq k_1} f(i,j)f(i,k)f(i_1,j)f(i_1,k_1)\bigg\}\\
&\quad + 256\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq j\neq k\neq k_1} f(i,j)f(i,k)f(i,j)f(i,k_1)\bigg\}\\
&\quad + 256\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq j\neq k\neq k_1} f(i,j)f(i,k)f(j,i)f(j,k_1)\bigg\}\\
&\quad + 512\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq i_1\neq j\neq k} f(i,j)f(i,k)f(i_1,i)f(i_1,j)\bigg\}\\
&\quad + 128\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq i_1\neq j\neq k} f(i,j)f(i,k)f(i_1,j)f(i_1,k)\bigg\}\\
&\quad + 128\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq j\neq k} f(i,j)f(i,k)f(i,j)f(i,k)\bigg\}\\
&\quad + 256\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1}E\bigg\{\sum_{i\neq j\neq k} f(i,j)f(i,k)f(j,i)f(j,k)\bigg\},
\end{aligned}$$

$$\equiv S_1 + S_2 + S_3 + S_4 + S_5 + S_6 + S_7 + S_8 + S_9 + S_{10}. \tag{3.44}$$

Under the null hypothesis, S_1, S_2, S_3, and S_4 are all zero. Terms S_9 and S_{10} are of the same order as R_2. Thus, $S_9 = o(I_1)$ and $S_{10} = o(I_1)$. Terms S_5, S_6, S_7, and S_8 are all of the same order in terms of n as term R_1. Additionally, using two iterations of the Cauchy Schwarz inequality, these terms will be the same as R_3 in terms of n and p. Thus, S_5, S_6, S_7, and S_8

are all smaller order terms in comparison to I_1. As a result, for some constant C,

$$I_2 \leq C\frac{([T\eta] - [T\nu])^2}{T^2}. \tag{3.45}$$

Finally, we show the order of I_3. Consider the possible indices for expanding

$$\Big\{ \sum_{i \neq j \neq k \neq l} f(i,j)f(k,l) \Big\}^2 = \sum_{i \neq j \neq k \neq l} \sum_{i_1 \neq j_1 \neq k_1 \neq l_1} f(i,j)f(k,l)f(i_1,j_1)f(k_1,l_1). \tag{3.46}$$

Let $F_c = \{i,j,k,l\} \cup \{i_1,j_1,k_1,l_1\}$ be the set of indices that are not equivalent to each other, where c represents the number of indices that are equivalent to each other in two sets $\{i,j,k,l\}$ and $\{i_1,j_1,k_1,l_1\}$. If there are no equivalent indices, then

$$F_0 = \{(i,j,k,l,i_1,j_1,k_1,l_1)\}.$$

Hence, the summation over F_0 is given by

$$\sum_{i \neq i_1 \neq j \neq j_1 \neq k \neq k_1 \neq l \neq l_1} f(i,j)f(k,l)f(i_1,j_1)f(k_1,l_1). \tag{3.47}$$

If there is one equivalent index, then

$$\begin{aligned} F_1 = \{&(i = i_1, j, k, l, j_1, k_1, l_1), (i = j_1, j, k, l, i_1, k_1, l_1), (i = k_1, j, k, l, i_1, j_1, l_1), \\ &(i = l_1, j, k, l, i_1, j_1, k_1), (i, j = i_1, k, l, j_1, k_1, l_1), (i, j = j_1, k, l, i_1, k_1, l_1), \\ &(i, j = k_1, k, l, i_1, j_1, l_1), (i, j = l_1, k, l, i_1, j_1, k_1)(i, j, k = i_1, l, j_1, k_1, l_1), \\ &(i, j, k = j_1, l, i_1, k_1, l_1), (i, j, k = k_1, l, i_1, j_1, l_1), (i, j, k = l_1, l, i_1, j_1, k_1) \\ &(i, j, k, l = i_1, j_1, k_1, l_1), (i, j, k, l = j_1, i_1, k_1, l_1), (i, j, k, l = k_1, i_1, j_1, l_1), \\ &(i, j, k, l = l_1, i_1, j_1, k_1)\}. \end{aligned}$$

The summation over F_1 is equivalent to

$$\sum_{i \neq j \neq j_1 \neq k \neq k_1 \neq l \neq l_1} 16 f(i,j)f(k,l)f(i,j_1)f(k_1,l_1). \tag{3.48}$$

If there are two equivalent indices, then

$$F_2 = \{(i = i_1, j = j_1, k, l, k_1, l_1), (i = i_1, j = k_1, k, l, j_1, l_1), (i = i_1, j = l_1, k, l, j_1, k_1),$$

$$(i = j_1, j = i_1, k, l, k_1, l_1), (i = j_1, j = k_1, k, l, i_1, l_1), (i = j_1, j = l_1, k, l, i_1, k_1),$$

$$(i = k_1, j = i_1, k, l, j_1, l_1), (i = k_1, j = j_1, k, l, i_1, l_1), (i = k_1, j = l_1, k, l, i_1, j_1),$$

$$(i = l_1, j = i_1, k, l, j_1, k_1), (i = l_1, j = j_1, k, l, i_1, k_1), (i = l_1, j = k_1, k, l, i_1, j_1),$$

$$(i = i_1, k = j_1, j, l, k_1, l_1), (i = i_1, k = k_1, j, l, j_1, l_1), (i = i_1, k = l_1, j, l, j_1, k_1),$$

$$(i = j_1, k = i_1, j, l, k_1, l_1), (i = j_1, k = k_1, j, l, i_1, l_1), (i = j_1, k = l_1, j, l, i_1, k_1),$$

$$(i = k_1, k = i_1, j, l, j_1, l_1), (i = k_1, k = j_1, j, l, i_1, l_1), (i = k_1, k = l_1, j, l, i_1, j_1),$$

$$(i = l_1, k = i_1, j, l, j_1, k_1), (i = l_1, k = j_1, j, l, i_1, k_1), (i = l_1, k = k_1, j, l, i_1, j_1),$$

$$(i = i_1, l = j_1, j, k, k_1, l_1), (i = i_1, l = k_1, j, k, j_1, l_1), (i = i_1, l = l_1, j, k, j_1, k_1),$$

$$(i = j_1, l = i_1, j, k, k_1, l_1), (i = j_1, l = k_1, j, k, i_1, l_1), (i = j_1, l = l_1, j, k, i_1, k_1),$$

$$(i = k_1, l = i_1, j, k, j_1, l_1), (i = k_1, l = j_1, j, k, i_1, l_1), (i = k_1, l = l_1, j, k, i_1, j_1),$$

$$(i = l_1, l = i_1, j, k, j_1, k_1), (i = l_1, l = j_1, j, k, i_1, k_1), (i = l_1, l = k_1, j, k, i_1, j_1),$$

$$(j = i_1, k = j_1, i, l, k_1, l_1), (j = i_1, k = k_1, i, l, j_1, l_1), (j = i_1, k = l_1, i, l, j_1, k_1),$$

$$(j = j_1, k = i_1, i, l, k_1, l_1), (j = j_1, k = k_1, i, l, i_1, l_1), (j = j_1, k = l_1, i, l, i_1, k_1),$$

$$(j = k_1, k = i_1, i, l, j_1, l_1), (j = k_1, k = j_1, i, l, i_1, l_1), (j = k_1, k = l_1, i, l, i_1, j_1),$$

$$(j = l_1, k = i_1, i, l, j_1, k_1), (j = l_1, k = j_1, i, l, i_1, k_1), (j = l_1, k = k_1, i, l, i_1, j_1),$$

$$(j = i_1, l = j_1, i, k, k_1, l_1), (j = i_1, l = k_1, i, k, j_1, l_1), (j = i_1, l = l_1, i, k, j_1, k_1),$$

$$(j = j_1, l = i_1, i, k, k_1, l_1), (j = j_1, l = k_1, i, k, i_1, l_1), (j = j_1, l = l_1, i, k, i_1, k_1),$$

$$(j = k_1, l = i_1, i, k, j_1, l_1), (j = k_1, l = j_1, i, k, i_1, l_1), (j = k_1, l = l_1, i, k, i_1, j_1),$$

$$(j = l_1, l = i_1, i, k, j_1, k_1), (j = l_1, l = j_1, i, k, i_1, k_1), (j = l_1, l = k_1, i, k, i_1, j_1),$$

$$(k = i_1, l = j_1, i, j, k_1, l_1), (k = i_1, l = k_1, i, j, j_1, l_1), (k = i_1, l = l_1, i, j, j_1, k_1),$$

$$(k = j_1, l = i_1, i, j, k_1, l_1), (k = j_1, l = k_1, i, j, i_1, l_1), (k = j_1, l = l_1, i, j, i_1, k_1),$$

$$(k = k_1, l = i_1, i, j, j_1, l_1), (k = k_1, l = j_1, i, j, i_1, l_1), (k = k_1, l = l_1, i, j, i_1, j_1),$$

$$(k = l_1, l = i_1, i, j, j_1, k_1), (k = l_1, l = j_1, i, j, i_1, k_1), (k = l_1, l = k_1, i, j, i_1, j_1)\}.$$

The summation over F_2 is equivalent to

$$\begin{aligned}
&\sum_{i\neq j\neq k\neq k_1\neq l\neq l_1} 4f(i,j)f(k,l)f(i,j)f(k_1,l_1)\\
&\quad+\sum_{i\neq j\neq j_1\neq k\neq l\neq l_1} 28f(i,j)f(k,l)f(i,j_1)f(j,l_1)\\
&\quad+\sum_{i\neq j\neq j_1\neq k\neq l\neq l_1} 32f(i,j)f(k,l)f(i,j_1)f(k,l_1).
\end{aligned}\tag{3.49}$$

If there are three equivalent indices, then

$$\begin{aligned}
F_3=\{&(i=i_1,j=j_1,k=k_1,l,l_1),(i=i_1,j=j_1,k=l_1,l,k_1),(i=i_1,j=k_1,k=j_1,l,l_1),\\
&(i=i_1,j=k_1,k=l_1,l,j_1),(i=i_1,j=l_1,k=j_1,l,k_1),(i=i_1,j=l_1,k=k_1,l,j_1),\\
&(i=j_1,j=i_1,k=k_1,l,l_1),(i=j_1,j=i_1,k=l_1,l,k_1),(i=j_1,j=k_1,k=i_1,l,l_1),\\
&(i=j_1,j=k_1,k=l_1,l,i_1),(i=j_1,j=l_1,k=i_1,l,k_1),(i=j_1,j=l_1,k=k_1,l,i_1),\\
&(i=k_1,j=i_1,k=j_1,l,l_1),(i=k_1,j=i_1,k=l_1,l,j_1),(i=k_1,j=j_1,k=i_1,l,l_1),\\
&(i=k_1,j=j_1,k=l_1,l,i_1),(i=k_1,j=l_1,k=i_1,l,j_1),(i=k_1,j=l_1,k=j_1,l,i_1),\\
&(i=l_1,j=i_1,k=j_1,l,k_1),(i=l_1,j=i_1,k=k_1,l,j_1),(i=l_1,j=j_1,k=i_1,l,k_1),\\
&(i=l_1,j=j_1,k=k_1,l,i_1),(i=l_1,j=k_1,k=i_1,l,j_1),(i=l_1,j=k_1,k=j_1,l,i_1),\\
&(i=i_1,j=j_1,l=k_1,k,l_1),(i=i_1,j=j_1,l=l_1,k,k_1),(i=i_1,j=k_1,l=j_1,k,l_1),\\
&(i=i_1,j=k_1,l=l_1,k,j_1),(i=i_1,j=l_1,l=j_1,k,k_1),(i=i_1,j=l_1,l=k_1,k,j_1),\\
&(i=j_1,j=i_1,l=k_1,k,l_1),(i=j_1,j=i_1,l=l_1,k,k_1),(i=j_1,j=k_1,l=i_1,k,l_1),\\
&(i=j_1,j=k_1,l=l_1,k,i_1),(i=j_1,j=l_1,l=i_1,k,k_1),(i=j_1,j=l_1,l=k_1,k,i_1),\\
&(i=k_1,j=i_1,l=j_1,k,l_1),(i=k_1,j=i_1,l=l_1,k,j_1),(i=k_1,j=j_1,l=i_1,k,l_1),\\
&(i=k_1,j=j_1,l=l_1,k,i_1),(i=k_1,j=l_1,l=i_1,k,j_1),(i=k_1,j=l_1,l=j_1,k,i_1),\\
&(i=l_1,j=i_1,l=j_1,k,k_1),(i=l_1,j=i_1,l=k_1,k,j_1),(i=l_1,j=j_1,l=i_1,k,k_1),\\
&(i=l_1,j=j_1,l=k_1,k,i_1),(i=l_1,j=k_1,l=i_1,k,j_1),(i=l_1,j=k_1,l=j_1,k,i_1),\\
&(i=i_1,k=j_1,l=k_1,j,l_1),(i=i_1,k=j_1,l=l_1,j,k_1),(i=i_1,k=k_1,l=j_1,j,l_1),\\
&(i=i_1,k=k_1,l=l_1,j,j_1),(i=i_1,k=l_1,l=j_1,j,k_1),(i=i_1,k=l_1,l=k_1,j,j_1),
\end{aligned}$$

$$
\begin{aligned}
&(i=j_1,k=i_1,l=k_1,j,l_1),(i=j_1,k=i_1,l=l_1,j,k_1),(i=j_1,k=k_1,l=i_1,j,l_1),\\
&(i=j_1,k=k_1,l=l_1,j,i_1),(i=j_1,k=l_1,l=i_1,j,k_1),(i=j_1,k=l_1,l=k_1,j,i_1),\\
&(i=k_1,k=i_1,l=j_1,j,l_1),(i=k_1,k=i_1,l=l_1,j,j_1),(i=k_1,k=j_1,l=i_1,j,l_1),\\
&(i=k_1,k=j_1,l=l_1,j,i_1),(i=k_1,k=l_1,l=i_1,j,j_1),(i=k_1,k=l_1,l=j_1,j,i_1),\\
&(i=l_1,k=i_1,l=j_1,j,k_1),(i=l_1,k=i_1,l=l_1,j,j_1),(i=l_1,k=j_1,l=i_1,j,k_1),\\
&(i=l_1,k=j_1,l=l_1,j,i_1),(i=l_1,k=k_1,l=i_1,j,j_1),(i=l_1,k=k_1,l=j_1,j,i_1),\\
&(j=i_1,k=j_1,l=k_1,i,l_1),(j=i_1,k=j_1,l=l_1,i,k_1),(j=i_1,k=k_1,l=j_1,i,l_1),\\
&(j=i_1,k=k_1,l=l_1,i,j_1),(j=i_1,k=l_1,l=j_1,i,k_1),(j=i_1,k=l_1,l=k_1,i,j_1),\\
&(j=j_1,k=i_1,l=k_1,i,l_1),(j=j_1,k=i_1,l=l_1,i,k_1),(j=j_1,k=k_1,l=i_1,i,l_1),\\
&(j=j_1,k=k_1,l=l_1,i,i_1),(j=j_1,k=l_1,l=i_1,i,k_1),(j=j_1,k=l_1,l=k_1,i,i_1),\\
&(j=k_1,k=i_1,l=j_1,i,l_1),(j=k_1,k=i_1,l=l_1,i,j_1),(j=k_1,k=j_1,l=i_1,i,l_1),\\
&(j=k_1,k=j_1,l=l_1,i,i_1),(j=k_1,k=l_1,l=i_1,i,j_1),(j=k_1,k=l_1,l=j_1,i,i_1),\\
&(j=l_1,k=i_1,l=j_1,i,k_1),(j=l_1,k=i_1,l=k_1,i,j_1),(j=l_1,k=j_1,l=i_1,i,k_1),\\
&(j=l_1,k=j_1,l=k_1,i,i_1),(j=l_1,k=k_1,l=i_1,i,j_1),(j=l_1,k=k_1,l=j_1,i,i_1)\}.
\end{aligned}
$$

The summation over F_3 is equivalent to

$$\sum_{i\neq j\neq k\neq l\neq l_1} 32f(i,j)f(k,l)f(i,j)f(k,l_1)+\sum_{i\neq j\neq k\neq l\neq l_1} 64f(i,j)f(k,l)f(i,k)f(j,l_1). \tag{3.50}$$

Lastly, if there are four equivalent indices, then

$$
\begin{aligned}
F_4=\{&(i=i_1,j=j_1,k=k_1,l=l_1),(i=i_1,j=j_1,k=l_1,l=k_1),\\
&(i=i_1,j=k_1,k=j_1,l=l_1),(i=i_1,j=k_1,k=l_1,l=j_1),\\
&(i=i_1,j=l_1,k=j_1,l=k_1),(i=i_1,j=l_1,k=k_1,l=j_1),\\
&(i=j_1,j=i_1,k=k_1,l=l_1),(i=j_1,j=i_1,k=l_1,l=k_1),\\
&(i=j_1,j=k_1,k=i_1,l=l_1),(i=j_1,j=k_1,k=l_1,l=i_1),\\
&(i=j_1,j=l_1,k=i_1,l=k_1),(i=j_1,j=l_1,k=k_1,l=i_1),
\end{aligned}
$$

$$(i=k_1,j=i_1,k=j_1,l=l_1),(i=k_1,j=i_1,k=l_1,l=j_1),$$

$$(i=k_1,j=j_1,k=i_1,l=l_1),(i=k_1,j=j_1,k=l_1,l=i_1),$$

$$(i=k_1,j=l_1,k=i_1,l=j_1),(i=k_1,j=l_1,k=j_1,l=i_1),$$

$$(i=l_1,j=i_1,k=j_1,l=k_1),(i=l_1,j=i_1,k=k_1,l=j_1),$$

$$(i=l_1,j=j_1,k=i_1,l=k_1),(i=l_1,j=j_1,k=k_1,l=i_1),$$

$$(i=l_1,j=k_1,k=i_1,l=j_1),(i=l_1,j=k_1,k=j_1,l=i_1)\}.$$

The summation over F_4 is equivalent to

$$\sum_{i\neq j\neq k\neq l} 6f(i,j)f(k,l)f(i,j)f(k,l)+\sum_{i\neq j\neq k\neq l} 18f(i,j)f(k,l)f(i,k)f(j,l). \tag{3.51}$$

As a result from (3.47) – (3.51),

$$\begin{aligned}
E\Big[\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)\Big\}^2\Big] &= E\Big\{\sum_{i\neq i_1\neq j\neq j_1\neq k\neq k_1\neq l\neq l_1} f(i,j)f(k,l)f(i_1,j_1)f(k_1,l_1)\Big\}\\
&+16E\Big\{\sum_{i\neq j\neq j_1\neq k\neq k_1\neq l\neq l_1} f(i,j)f(k,l)f(i,j_1)f(k_1,l_1)\Big\}\\
&+4E\Big\{\sum_{i\neq j\neq k\neq k_1\neq l\neq l_1} f(i,j)f(k,l)f(i,j)f(k_1,l_1)\Big\}\\
&+28E\Big\{\sum_{i\neq j\neq j_1\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,j_1)f(j,l_1)\Big\}\\
&+32E\Big\{\sum_{i\neq j\neq j_1\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,j_1)f(k,l_1)\Big\}\\
&+32E\Big\{\sum_{i\neq j\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,j)f(k,l_1)\Big\}\\
&+64E\Big\{\sum_{i\neq j\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,k)f(j,l_1)\Big\}\\
&+6E\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)f(i,j)f(k,l)\Big\}\\
&+18E\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)f(i,k)f(j,l)\Big\}
\end{aligned}$$

Thus,

$$I_3 = 2\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big[\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)\Big\}^2\Big],$$

$$= 2\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq i_1\neq j\neq j_1\neq k\neq k_1\neq l\neq l_1} f(i,j)f(k,l)f(i_1,j_1)f(k_1,l_1)\Big\}$$

$$+ 32\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq j_1\neq k\neq k_1\neq l\neq l_1} f(i,j)f(k,l)f(i,j_1)f(k_1,l_1)\Big\}$$

$$+ 8\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq k\neq k_1\neq l\neq l_1} f(i,j)f(k,l)f(i,j)f(k_1,l_1)\Big\}$$

$$+ 56\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq j_1\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,j_1)f(j,l_1)\Big\}$$

$$+ 64\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq j_1\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,j_1)f(k,l_1)\Big\}$$

$$+ 64\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,j)f(k,l_1)\Big\}$$

$$+ 128\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq k\neq l\neq l_1} f(i,j)f(k,l)f(i,k)f(j,l_1)\Big\}$$

$$+ 12\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)f(i,j)f(k,l)\Big\}$$

$$+ 36\{n^2(n-1)^2 tr^4(\Sigma^2)T^6\}^{-1} E\Big\{\sum_{i\neq j\neq k\neq l} f(i,j)f(k,l)f(i,k)f(j,l)\Big\},$$

$$\equiv Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_6 + Q_7 + Q_8 + Q_9. \tag{3.52}$$

Due to the mutually different indices, Q_1, Q_2, Q_3, Q_4 all equal zero since $E\Big\{f(i,j)\Big\} = 0$ for i different than j. Furthermore, under the null hypothesis, Q_5, Q_6, and Q_7 all equal zero. For Q_5, consider $E\Big\{f(i,j)f(k,l)f(i,j_1)f(k,l_1)\Big\}$. Due to the mutually different indices $E\Big\{f(i,j)f(k,l)f(i,j_1)f(k,l_1)\Big\} = E\Big\{f(i,j)f(i,j_1)\Big\}E\Big\{f(k,l)f(k,l_1)\Big\}$. By (3.25), each of the expectation terms is zero and thus $Q_5 = 0$. Similarly, for

Q_6, $E\Big\{f(i,j)f(k,l)f(i,j)f(k,l_1)\Big\} = E\Big\{f(i,j)f(i,j)\Big\}E\Big\{f(k,l)f(k,l_1)\Big\}$. Again, by (3.25) the term $E\Big\{f(k,l)f(k,l_1)\Big\}$ is zero. Thus, $Q_6 = 0$. To see that Q_7 is zero, consider

$$\begin{aligned} E\Big\{f(i,j)f(k,l)f(i,k)f(j,l_1)\Big\} = E\Big\{&\Big(\sum_{\mathcal{S}^1}\tilde{W}_{s_1s_2}(i,j) - \sum_{\mathcal{S}^2}\tilde{W}_{s_1s_2}(i,j)\Big)\\ &\times\Big(\sum_{\mathcal{R}^1}\tilde{W}_{r_1r_2}(k,l) - \sum_{\mathcal{R}^2}\tilde{W}_{r_1r_2}(k,l)\Big)\\ &\times\Big(\sum_{\mathcal{U}^1}\tilde{W}_{u_1u_2}(i,k) - \sum_{\mathcal{U}^2}\tilde{W}_{u_1u_2}(i,k)\Big)\\ &\times\Big(\sum_{\mathcal{X}^1}\tilde{W}_{x_1x_2}(j,l_1) - \sum_{\mathcal{X}^2}\tilde{W}_{x_1x_2}(j,l_1)\Big)\Big\}, \end{aligned}$$

$$\begin{aligned} &= \sum_{a,b,c,d=1}^{2}(-1)^{|a-b|+|c-d|}\sum_{\mathcal{S}^a}\sum_{\mathcal{R}^b}\sum_{\mathcal{U}^c}\sum_{\mathcal{X}^d}\\ &\quad\times E\Big\{\tilde{W}_{s_1s_2}(i,j)\tilde{W}_{r_1r_2}(k,l)\tilde{W}_{u_1u_2}(i,k)\tilde{W}_{x_1x_2}(j,l_1)\Big\}. \end{aligned}$$

Accordingly, $E\Big\{\tilde{W}_{s_1s_2}(i,j)\tilde{W}_{r_1r_2}(k,l)\tilde{W}_{u_1u_2}(i,k)\tilde{W}_{x_1x_2}(j,l_1)\Big\}$ can be expressed as

$$\sum_{a,b,c,d,e,f,g,h=1}^{2}(-1)^{|a-b|+|c-d|+|e-f|+|g-h|}E\Big\{V_{s_as_b}(i,j)V_{r_cr_d}(k,l)V_{u_eu_f}(i,k)V_{x_gx_h}(j,l_1)\Big\}.$$

Under the null hypothesis,

$$\begin{aligned} E\Big\{V_{s_as_b}(i,j)V_{r_cr_d}(k,l)V_{u_eu_f}(i,k)V_{x_gx_h}(j,l_1)\Big\} &= \mathrm{tr}(\Sigma^4) + 2tr^2(\Sigma^2)\mathrm{tr}(\Sigma C_{s_bx_g}\Sigma C_{x_gs_b})\\ &+ 2tr^2(\Sigma^2)\mathrm{tr}(\Sigma C_{s_au_e}\Sigma C_{u_es_a}) + 2tr^2(\Sigma^2)\mathrm{tr}(\Sigma C_{r_cu_f}\Sigma C_{u_fr_c})\\ &+ 4tr(\Sigma^2)\mathrm{tr}(\Sigma C_{u_es_a}C_{s_bx_g}\Sigma C_{x_gs_b}C_{s_au_e}) + 4tr(\Sigma^2)\mathrm{tr}(\Sigma C_{r_cu_f}C_{u_es_a}\Sigma C_{s_au_e}C_{u_fr_c})\\ &+ 4\mathrm{tr}(\Sigma C_{r_cu_f}\Sigma C_{u_fr_c})\mathrm{tr}(\Sigma C_{s_bx_g}\Sigma C_{x_gs_b})\\ &+ 8tr(\Sigma C_{r_cu_f}C_{u_es_a}C_{s_bx_g}\Sigma C_{x_gs_b}C_{s_au_e}C_{u_fr_c}). \end{aligned}$$

The summation of the above expression over $a,b,c,d,e,f,g,h \in \{1,2\}$ is zero. Hence, $Q_7 = 0$. Terms Q_8 and Q_9 are at most the order of R_1. Term Q_8 has the same order up to a constant as R_1 due to the four mutually different indices. By the Cauchy-Schwarz inequality, the $E\Big\{f(i,j)f(k,l)f(i,k)f(j,l)\Big\}$ with respect to Q_9 can be expressed as

$$\begin{aligned}
E\Big\{f(i,j)f(k,l)f(i,k)f(j,l)\Big\} \le{}& \Big[E\Big\{f(i,j)f(k,l)f(i,j)f(k,l)\Big\}\Big]^{1/2} \\
&\times \Big[E\Big\{f(i,k)f(j,l)f(i,k)f(j,l)\Big\}\Big]^{1/2}, \\
\\
\asymp{}& \mathcal{O}\Big(E\Big\{f(i,j)f(i,j)f(k,l)f(k,l)\Big\}\Big).
\end{aligned}$$

Therefore,

$$Q_9 \le Q_8 \asymp R_1 \asymp C\frac{([T\eta]-[T\nu])^2}{T^2}.$$

As a result,

$$I_3 \le C\frac{([T\eta]-[T\nu])^2}{T^2}, \tag{3.53}$$

for some constant C. In summary, $E[\{G_n(\eta)-G_n(\nu)\}^4] \le C([T\eta]-[T\nu])^2/T^2$.

Let $0 \le i \le j \le T$. By the definitions of η and ν, (3.37), (3.45), (3.53), and Markov's inequality, it follows that for any $\lambda > 0$,

$$\begin{aligned}
\mathrm{pr}\Big(|G_n(i/T)-G_n(j/T)| \ge \lambda\Big) &\le \frac{E\Big\{|G_n(i/T)-G_n(j/T)|^4\Big\}}{\lambda^4} \\
&\le C\left[\frac{(j-i)}{\lambda^2 T}\right]^2 \\
&= \frac{C}{\lambda^{4\beta}}\left(\frac{1}{T}\sum_{i<l\le j} u_l\right)^{2\alpha},
\end{aligned}$$

where we set $\alpha = 1$, $\beta = 1$, and $u_l = l-(l-1) = 1$. By Theorem 10.2 in Billingsley (1999)

$$\begin{aligned}
\mathrm{pr}\Big\{\max_{t\in\mathcal{T}}|G_n(i/T)| \ge \lambda\Big\} &\le \frac{KC}{\lambda^{4\beta}}\left(\frac{1}{T}\sum_{0<l\le T} u_l\right)^{2\alpha}, \\
&= \frac{KC}{\lambda^4}
\end{aligned}$$

For λ large, above probability is less than any $\varepsilon > 0$. As a result, $\max_{t\in\mathcal{T}}|G_n(i/T)|$ is tight, and thus $\max_{t\in\mathcal{T}}\sigma_{nt,0}^{-1}\hat{D}_{nt}$ is also tight. Therefore, by the tightness of the stochastic process and convergence of the finite dimensional distributions, it follows that $\mathcal{M}_n$ converges to a Gaussian process with mean zero and correlation R_z. $\square$

Proof of Theorem 9. Assume that one change point exists at time τ. Thus, we assume alternative H_1^* as defined in (3.15). Let $\Delta_p = \mathrm{tr}\{(\Sigma_1 - \Sigma_T)^2\}$ and $\nu_{t,\max} = \max_{t\in\mathcal{T}}\max\left(\sqrt{V_{0t}/w^2(t)}, \sqrt{nV_{1t}/w^2(t)}\right)$ such that $\mathcal{T} = \{1,\dots,T-1\}$.

Define a set, $E(C)$, such that $E(C) = \{t \in \{1,\dots,T-1\} : |t-\tau| \geq C\Theta\}$, where C is some constant and Θ is a function of p, n, and T. The value Θ is chosen to show the rate of convergence of the change point estimator under the asymptotic setting where p, n, and T diverge. Thus, to establish this rate of convergence we must show that for some C, $\mathrm{pr}(|\hat{\tau}-\tau| \geq C\Theta) < \varepsilon$. It is sufficient to show that $\mathrm{pr}(\max_{t\in E(C)}\hat{D}_{nt} > \hat{D}_{n\tau}) < \varepsilon$ since $\mathrm{pr}(|\hat{\tau}-\tau| \geq C\Theta) = \mathrm{pr}(\hat{\tau} \in E(C)) \leq \mathrm{pr}(\max_{t\in E(C)}\hat{D}_{nt} > \hat{D}_{n\tau})$ for $\{\hat{\tau} \in E(C)\} \subset \{\max_{t\in E(C)}\hat{D}_{nt} > \hat{D}_{n\tau}\}$. Thus,

$$\begin{aligned}
\mathrm{pr}\Big(\max_{t\in E(C)}\hat{D}_{nt} > \hat{D}_{n\tau}\Big) &\leq \sum_{t\in E(C)}\mathrm{pr}\Big(\hat{D}_{nt} > \hat{D}_{n\tau}\Big)\\
&= \sum_{t\in E(C)}\mathrm{pr}\Big(\hat{D}_{nt} - D_t + D_t - D_\tau > \hat{D}_{n\tau} - D_\tau\Big)\\
&= \sum_{t\in E(C)}\mathrm{pr}\Big[\{\hat{D}_{nt} - D_t\} + \{-(\hat{D}_{n\tau} - D_\tau)\} > -\{D_t - D_\tau\}\Big]\\
&\leq \sum_{t\in E(C)}\mathrm{pr}\Big[|\{\hat{D}_{nt} - D_t\} + \{-(\hat{D}_{n\tau} - D_\tau)\}| > -\{D_t - D_\tau\}\Big].
\end{aligned}$$

The term $-(D_t - D_\tau)$ can be expressed as $|t-\tau|G(t;\tau)\Delta_p$, where

$$G(t;\tau) = \begin{cases} \dfrac{1}{T-t}, & 1 \leq t \leq \tau,\\[2em] \dfrac{1}{t}, & \tau+1 \leq t < T. \end{cases}$$

In terms of T, the function G is of the order $1/T$.

Recall that for two random variables, X and Y,

$$\mathrm{pr}(|X+Y|>\varepsilon)\le \mathrm{pr}(|X|>\varepsilon/2)+\mathrm{pr}(|Y|>\varepsilon/2).$$

Hence,

$$\begin{aligned}
\mathrm{pr}\Big(\max_{t\in E(C)}\hat{D}_{nt}>\hat{D}_{n\tau}\Big) &\le \sum_{t\in E(C)}\mathrm{pr}\Big[|\{\hat{D}_{nt}-D_t\}+\{-(\hat{D}_{n\tau}-D_\tau)\}|>-\{D_t-D_\tau\}\Big]\\
&\le \sum_{t\in E(C)}\mathrm{pr}\bigg\{|\hat{D}_{nt}-D_t|>\frac{|t-\tau|G(t;\tau)\Delta_p}{2}\bigg\}\\
&\quad+\sum_{t\in E(C)}\mathrm{pr}\bigg\{|\hat{D}_{n\tau}-D_\tau|>\frac{|t-\tau|G(t;\tau)\Delta_p}{2}\bigg\}\\
&=\sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{nt}-D_t|}{\sigma_{nt}}>\frac{|t-\tau|G(t;\tau)\Delta_p}{2\sigma_{nt}}\bigg\}\\
&\quad+\sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{n\tau}-D_\tau|}{\sigma_{n\tau}}>\frac{|t-\tau|G(t;\tau)\Delta_p}{2\sigma_{n\tau}}\bigg\}\\
&=\sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{nt}-D_t|}{\sigma_{nt}}>\frac{|t-\tau|G(t;\tau)n\Delta_p}{2\sqrt{4\tilde{V}_{0t}+8n\tilde{V}_{1t}}}\bigg\}\\
&\quad+\sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{n\tau}-D_\tau|}{\sigma_{n\tau}}>\frac{|t-\tau|G(t;\tau)n\Delta_p}{2\sqrt{4\tilde{V}_{0\tau}+8n\tilde{V}_{1\tau}}}\bigg\}\\
&\le \sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{nt}-D_t|}{\sigma_{nt}}>\frac{C_1|t-\tau|G(t;\tau)n\Delta_p}{2\nu_{t,\max}}\bigg\}\\
&\quad+\sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{n\tau}-D_\tau|}{\sigma_{n\tau}}>\frac{C_1|t-\tau|G(t;\tau)n\Delta_p}{2\nu_{t,\max}}\bigg\}\\
&\le \sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{nt}-D_t|}{\sigma_{nt}}>\frac{C\Theta G(t;\tau)n\Delta_p}{\nu_{t,\max}}\bigg\}\\
&\quad+\sum_{t\in E(C)}\mathrm{pr}\bigg\{\frac{|\hat{D}_{n\tau}-D_\tau|}{\sigma_{n\tau}}>\frac{C\Theta G(t;\tau)n\Delta_p}{\nu_{t,\max}}\bigg\},
\end{aligned}$$

for some constants C_1 and C. Choose $\Theta=\nu_{t,\max}T\sqrt{\log T}/n\Delta_p$. By the choice of Θ, order of $G(t;\tau)$, and the fact that both $(\hat{D}_{nt} - D_t)/\sigma_{nt}, (\hat{D}_{n\tau} - D_\tau)/\sigma_{n\tau} \sim N(0,1)$, it follows that

$$\sum_{t \in E(C)} \mathrm{pr}\left\{ \frac{|\hat{D}_{nt} - D_t|}{\sigma_{nt}} > \frac{C\Theta G(t;\tau) n\Delta_p}{\nu_{t,\max}} \right\} \le \sum_{t \in E(C)} \mathrm{pr}\left(|Z| > \sqrt{C \log T} \right), \tag{3.54}$$

$$\sum_{t \in E(C)} \mathrm{pr}\left\{ \frac{|\hat{D}_{n\tau} - D_\tau|}{\sigma_{n\tau}} > \frac{C\Theta G(t;\tau) n\Delta_p}{\nu_{t,\max}} \right\} \le \sum_{t \in E(C)} \mathrm{pr}\left(|Z| > \sqrt{C \log T} \right), \tag{3.55}$$

where $Z \sim N(0,1)$ and C is some constant.

Recall that for a standard normal random variable, Z, and for any $k > 0$, $\mathrm{pr}(|Z| > k) \le 2\exp\{-x^2/2\}$. For a large enough C, the summation terms in (3.54) and (3.55) can be expressed as

$$\sum_{t \in E(C)} \mathrm{pr}\left(|Z| > \sqrt{C \log T} \right) \le \sum_{t \in E(C)} 2T^{-\frac{C}{2}} < \varepsilon.$$

For large C, the series is convergent as $T \to \infty$. Therefore, $\mathrm{pr}(\max_{t \in E(C)} \hat{D}_{nt} > \hat{D}_{n\tau}) < \varepsilon$, and

$$\hat{\tau} - \tau = O_p\left(\frac{\nu_{t,\max} T \sqrt{\log T}}{n\Delta_p} \right)$$

for $\Delta_p = \mathrm{tr}\{(\Sigma_1 - \Sigma_T)^2\}$ and $\nu_{t,\max} = \max_{t \in \mathcal{T}} \max\left(\sqrt{V_{0t}/w^2(t)}, \sqrt{nV_{1t}/w^2(t)} \right)$. The rate of convergence can be simplified further. The function $w^{-1}(t)$ is minimized at $T/2$. Therefore,

$$\hat{\tau} - \tau = O_p\left(\frac{\nu_{\max} \sqrt{\log T}}{n\Delta_p} \right)$$

for $\Delta_p = \mathrm{tr}\{(\Sigma_1 - \Sigma_T)^2\}$ and $\nu_{\max} = \max_{t \in \mathcal{T}} \max\left(\sqrt{V_{0t}}, \sqrt{nV_{1t}} \right)$. □

Proof of Theorem 10. Recall Theorem 5: Under the alternative H_1 of (3.1), the maximum value of D_t is attained at one of the q change points. We will make use of this theorem in the proof that follows.

We will first show that provided change points exist, we can detect their existence with probability one, and we can the locations of the change points with probability one. Assume at least one change point exists in the interval I_t and that the cardinality of $\hat{\mathcal{Q}}$ is less than the cardinality of $\mathcal{Q}$. To show we can detect the existence of change points with probability one in the interval I_t we must show that $\mathrm{pr}(\mathcal{M}_n[I_t] > W_{\alpha_n}[I_t]) = 1$.

$$\begin{aligned}
\mathrm{pr}(\mathcal{M}_n[I_t] > W_{\alpha_n}[I_t]) &\geq \mathrm{pr}\Big(\sigma_{nt,0}^{-1}[I_t]\hat{D}_{nt}[I_t] > W_{\alpha_n}[I_t]\Big) \\
&= 1 - \mathrm{pr}\Big(\sigma_{nt,0}^{-1}[I_t]\hat{D}_{nt}[I_t] \leq W_{\alpha_n}[I_t]\Big) \\
&= 1 - \mathrm{pr}\bigg(\frac{\hat{D}_{nt}[I_t] - D_t[I_t]}{\sigma_{nt}^{-1}[I_t]} \leq \frac{\sigma_{nt,0}[I_t]W_{\alpha_n}[I_t] - D_t[I_t]}{\sigma_{nt}^{-1}[I_t]}\bigg) \\
&= 1 - \mathrm{pr}\bigg(Z \leq \frac{\sigma_{nt,0}[I_t]}{\sigma_{nt}^{-1}[I_t]}W_{\alpha_n}[I_t] - \frac{D_t[I_t]}{\sigma_{nt}^{-1}[I_t]}\bigg) \\
&\to 1,
\end{aligned}$$

where Z is a standard normal random variable. The $\mathrm{pr}\bigg(Z \leq \frac{\sigma_{nt,0}[I_t]}{\sigma_{nt}^{-1}[I_t]}W_{\alpha_n}[I_t] - \frac{D_t[I_t]}{\sigma_{nt}^{-1}[I_t]}\bigg)$ goes to zero by our premise that $W_{\alpha_n} = o(\mathrm{mSNR})$ for any I_t. Therefore, it follows that we can detect the existence of a change point with probability one. Furthermore, by Theorem 5, Theorem 9 and our premise that $\nu_{\max}[I_t]\sqrt{\log(T)}/n\Delta_p[I_t] \to 0$, we can also correctly identify a change point with probability one. The above derivations do not depend on I_t since each subsequence satisfies the premises of this Theorem.

We also need to demonstrate that no change points will be identified that are not true change points. Thus, consider the case where $\hat{\mathcal{Q}} = \mathcal{Q}$. It is sufficient to demonstrate that no change point will be detected among the remaining time interval segments. Under H_0 of (3.1), as $\alpha_n \to \infty$, it follows that by Theorem 8 $\mathrm{pr}(\mathcal{M}_n[I_t] > W_{\alpha_n}[I_t]) = \alpha_n \to 0$ for some interval I_t with no change points. Therefore, no change points will be incorrectly-identified at any stage of the binary segmentation procedure. As a result $\hat{\mathcal{Q}} \to \mathcal{Q}$ in probability. $\square$

CHAPTER 4

A HIDDEN MARKOV APPROACH FOR QTL MAPPING USING ALLELE-SPECIFIC EXPRESSION SNPS

4.1 Introduction

Allele-specific expression (ASE) is part of the foundation for genetic diversity and is paramount to programming and development of biological cells (Ferguson-Smith 2001). ASE serves as a proxy for differential expression of two alleles at the same location within an organism (Gu & Wang 2015). For example, allele-specific expression can be characterized as the ratio between allele A and allele T. Differential expression is primarily explained by three factors: cis-acting modification, post transcription modification, and epigenetic modification (Ferguson-Smith 2001). Cis-effects correspond to the allele-specific variation, and thus, by quantifying ASE it is possible to identify cis-acting effects on an inter-individual basis among heterozygous individuals (Buckland 2004). The presence of ASE implies one or multiple variants have cis-acting effects on gene expression levels that could be directly correlated to phenotypic variation (Skelly et al. 2011). In fact, the phenomena of ASE has become a focal point in identifying predispositions towards certain diseases (de la Chapelle 2009). Due to the importance of understanding ASE, two natural questions arise with regards to its influence on phenotypic traits. What is the relationship between single nucleotide polymorphisms (SNPs) with ASE and a phenotypic trait? Which SNPs with ASE are have an effect on phenotypic variation? Our focus in Chapter 4 is to develop a procedure using a novel hierarchical model to answer the second question.

Quantitative trait loci (QTL) mapping is the statistical process of identifying locations in the genome that have an association with a complex phenotypic trait. For example, geneticists may be interested in understanding which genes affect cholesterol. Their understanding of this association can provide insight towards disease prevention and susceptibility. An effective QTL mapping procedure can also provide researchers a better understanding of breeding and appropriate techniques, and permit altered genetic variation within a population (Cheng et al. 2015). Studying SNPs with ASE and phenotypic variation was shown to be successful by Cheng et al. (2015). By applying multiple Bayesian approaches, Cheng et al. (2015) identified genetic markers in chickens associated with a resistance to Marek's disease. This disease is highly contagious and results in paralysis of the animal. The potential to eradicate Marek's disease through superior breeding techniques would be valuable to farmers and individuals within the animal science community. Cheng et al. (2015) discovered that 83% of the genetic variance in Marek's disease resistance was explained by the selected SNPs exhibiting ASE. These results were validated through a progeny study that found a 22% difference in the occurrence of Marek's disease after one generation of bidirectional selection (Cheng et al. 2015). The profound discovery by Cheng et al. (2015) gives credence to the fact that gene expression explains a large portion of phenotypic variation.

Next generation RNA sequencing data is now being widely used to investigate the presence of ASE. However, inference with regards to ASE remains a challenge along with mapping quantitative trait loci in the presence of only RNA sequencing data (Skelly et al. 2011). Skelly et al. (2011) proposed a three-stage hierarchical Bayesian model to test ASE gene expression and study cis-regulatory variation. However, their procedure requires genomic DNA data to establish prior probabilities. Similarly, Nariai et al. (2016) established a Bayesian framework with variational inference for estimating allele-specific expression. Their technique also relied on diploid DNA data and did not link any phenotypic response. Hu et al. (2015) proposed a unified maximum likelihood approach combining two models based on ASE and total RNA read counts. Their approach involved cis-expression QTL mapping with RNA sequencing data via a beta-binomial distribution.

In this chapter we present a novel two-step approach to perform QTL mapping using SNPs with allele-specific expression. In the first step, we predict the ASE ratios from RNA sequencing data. In step two, we use the predicted ASE ratios to identify SNPs with cis-acting effects as it relates to a phenotypic response variable. We elicit a hierarchical model for analysis of RNA sequence data to discover polymorphisms in expressed sequences whose allele-specific expression is correlated with observed phenotypic variation. In our hierarchical model, we first implement a hidden Markov approach to impute the underlying genotype and ASE status combinations based on the RNA read count data and simultaneously predict ASE ratios for heterozygous SNP locations. Second, we apply regularized regression to identify SNPs with ASE ratios that significantly impact an observed phenotypic response. Ordinary least squares is then applied for refinement. Our proposed hierarchical model and procedure has several advantages over existing methods. First, the hidden Markov model allows us to model dependence among SNPs and affords accurate genotype-ASE status imputation given RNA read counts (Steibel et al. 2015). Second, our procedure obtains an ASE ratio estimate in the absence of genetic DNA data. Many of the existing techniques required genetic DNA data for ASE estimation. Third, our proposed model integrates RNA sequencing data and phenotypic data to make inferences about the ASE status and cis-acting affects on phenotypic data. Fourth, our proposed method is easy to implement, where parameter estimation for the hidden Markov model is performed using the expectation-maximization (EM) algorithm. Variable selection via cyclic coordinate descent allows us to identify significant SNPs quickly and accurately, given an adequate signal-to-noise ratio. Lastly, our hierarchical model offers flexibility with regards to the phenotypic response model of interest, mapping error, spatial dependency, and individual variation in ASE ratios (Steibel et al. 2015).

Chapter 4 is organized as follows. In Section 4.2 we introduce the first layer of our proposed model. A hidden Markov model and genotype-ASE status with ASE prediction is recited based on the results of Steibel et al. (2015). In Section 4.3 we propose our method to identify SNPs with ASE that have cis-acting effects on a phenotypic variable of interest. Simulation results and a comparison with two competing procedures are detailed in Section 4.4. Our procedure is applied to a real data example that combines RNA sequencing data and phenotypic data from a sounder of swine in Section 4.5. The swim data set and a procedure to implement the hidden Markov approach is available in the R package HMMASE at `http://www.stt.msu.edu/users/pszhong/HMMASE.html`.

4.2 A hidden Markov model for SNP genotype calling

In this Section we introduce the basic setting and proposed model in Steibel et al. (2015). We introduce the salient features of their model, HMM-ASE, before we concentrate on ASE prediction and quantitative trait loci mapping in Section 4.3.

Let $X_{il} = (X_{il1}, X_{il2}, X_{il3}, X_{il4})^{\mathrm{T}}$ be a random vector of RNA read counts at the lth SNP for the ith individual. Denote x_{il} $(l = 1, \ldots, L;\ i = 1, \ldots, n)$ as the observed RNA read counts, where $x_{il1}, x_{il2}, x_{il3}, x_{il4}$ represent observed counts for alleles A, C, G, and T, respectively. Define the total RNA read counts at SNP l for individual i as $n_{il} = \sum_{j=1}^{4} x_{ilj}$.

Below we provide a set-up for a hidden Markov model with only two possible alleles: A or T. The procedure can easily be extended to consider a non-bi-allelic SNP. Let G_{il} be a latent variable that describes the genotype-ASE status with five possible hidden states. For each individual i at SNP l let

$$G_{il} = \begin{cases} 1 & \text{for “AA”}, \\ 2 & \text{for “AT-NASE”}, \\ 3 & \text{for “AT-ASE-HIGH”}, \\ 4 & \text{for “AT-ASE-LOW”}, \\ 5 & \text{for “TT”}. \end{cases} \tag{4.1}$$

Two homozygous hidden states are represented by AA and TT, and three heterozygous states are classified according to a relative ASE level. The variable G_{il} is latent and we assume that $G_{il} (l = 1, \ldots, L)$ follows a Markov process. Let A be the probability transition matrix for the Markov process G_{il}. Define the transition probabilities as

$$\mathrm{pr}(G_{il} = k' | G_{i(l-1)} = k) = a_{kk'} \qquad k, k' = 1, \ldots, 5. \tag{4.2}$$

Let π_{ik} $(i = 1, \ldots, n;\ k = 1, \ldots, 5)$ be the the initial probabilities of G_{il1} being a specific state in (4.1) such that $\mathrm{pr}(G_{i1} = k) = \pi_{ik}$.

We assume that the RNA read counts are generated by a hierarchical model conditional on the underlying state of G_{il} and ASE ratios. Let δ_{il} be a random variable for ASE ratios conditional on G_{il}. Thus,

$$\delta_{il}|G_{il} = k \sim \begin{cases} I_{\{\delta_{il}=1\}} & \text{for } k = 1, \\ I_{\{\delta_{il}=0.5\}} & \text{for } k = 2, \\ \text{Beta}_{(0.5,1]}(\alpha_1, \beta_1) & \text{for } k = 3, \\ \text{Beta}_{[0,0,5)}(\alpha_2, \beta_2) & \text{for } k = 4, \\ I_{\{\delta_{il}=0\}} & \text{for } k = 5. \end{cases} \tag{4.3}$$

If the underlying genotype is homozygous, then the corresponding ASE ratio is either zero or one with probability one. If the underlying genotype is heterozygous, but without ASE, then the corresponding ASE ratio is 0.5 with probability one. For the two remaining heterozygous states, it follows that the ASE ratio is defined as $\text{Beta}_{(0.5,1]}(\alpha_1, \beta_1)$ and $\text{Beta}_{[0,0,5)}(\alpha_2, \beta_2)$, where each represent scaled beta distributions with scale and shape parameters being α_1, α_2 and β_1, β_2, respectively. Conditional on G_{il}, we assume that δ_{il} are independent.

The first layer of the hierarchical model is conditional on a latent genotype-ASE status. In the second layer of the hierarchical model we define the probability distribution for RNA read counts conditional on 4.3. gives us the distribution of the RNA read counts. Here we assume that $X_{il} = (X_{il1}, X_{il2}, X_{il3}, X_{il4})^{\mathrm{T}}$ conditional on δ_{il} follows a multinomial distribution such that

$$X_{il}|\delta_{il} \sim Multinomial(n_{il}, p(\delta_{il}, e)), \tag{4.4}$$

where $p(\delta_{il}, e) = \left(\left(1 - \frac{4e}{3}\right)\delta_{il} + \frac{e}{3}\ , \frac{e}{3}\ , \frac{e}{3}\ , \left(\frac{4e}{3} - 1\right)\delta_{il} + 1 - e\right)$ is the probability vector in the multinomial distribution for A, C, G, and T, respectively. We assume that all reads are observable via a mapping error parameter denoted as e. If $e = 0$, then $p(\delta_{il}, 0) = (\delta_{il}, 0, 0, 1 - \delta_{il})$ represents the probabilities for observing A, C, G, and T, respectively.

Figure 4.1 illustrates our hidden Markov model specification for the ith individual when $L = 5$. The hidden variables G_{il} are dependent via a Markov process. The variables δ_{il} are conditional on G_{il} and independent among each other. RNA read counts are conditional on G_{il}, but through δ_{il}.

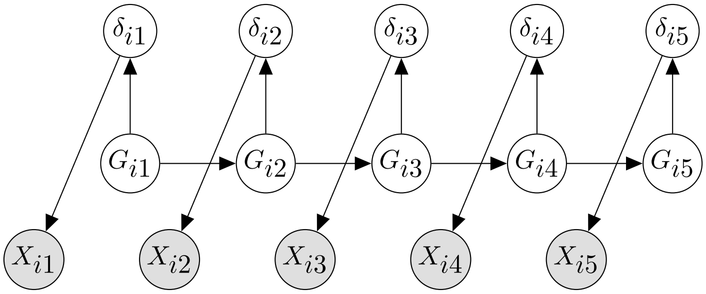

Figure 4.1: A graphical model for illustrating the hidden Markov model for SNP genotype calling. Grey circles represent observed values. White circles represent latent variables.

Given observed RNA read counts, x_{il}, we can predict the underlying genotype-ASE status, G_{il}, via the expectation-maximization (EM) algorithm and forward-backward procedure. In addition, and more importantly, given observed RNA read counts and underlying genotype-ASE statuses, we can derive the distribution for allele-specific expression ratios and use the posterior mode of the distribution as an estimate for the ratio of ASE.

4.3 Phenotypic model specification

Our ultimate goal is to identify significant SNPs and understand their affects on phenotypic variation. Let Y_i be a phenotypic response of interest, where

$$Y_i \sim f_{Y_i}(y_i|\tau_i, \phi) = \exp\left[\frac{y_i\tau_i - b(\tau_i)}{a(\phi)} + c(y_i, \phi)\right], \qquad i = 1, \ldots, n. \tag{4.5}$$

We assume the distribution of Y_i is in the form of a known exponential family. Let τ be the canonical parameter and let ϕ be the dispersion parameter. For example, suppose the phenotypic trait is eye color. If eye color is binary, such as the case for blue eyes or not blue eyes, then we assume (4.5) follows a Bernoulli distribution. However, if the phenotype is continuous one may consider the distribution of (4.5) to be Gaussian or Exponential. Furthermore, let $\eta(\delta_i) = \sum_{l=1}^{L} \delta_{il}\gamma_l$, where γ is an L-dimensional vector of unknown parameters that represent the effects of gene expression to the phenotypic response Y_i and δ_i is an L-

dimensional vector of ASE ratios. In order to relate the parameters of the distribution to the various predictors we denote $E[Y_i] = \mu_i$. Thus, for a canonical link function, h, it is the case that $\eta(\delta_i) = h(\mu_i) = \tau_i$. In particular, if we assume Y_i follows a normal distribution, then h is the identity link; whereas if Y_i follows a Bernoulli distribution, then h could be the logit link.

4.3.1 Prediction of ASE ratios

ASE ratios are unknown random variables. If we want to use them as predictors with regards to modeling a phenotypic response, then we need an estimation procedure. We consider two posterior probabilities that will be useful in our ultimate goal of identifying significant SNPs. Calculation of these two posterior distributions is dependent on an unknown parameter vector $\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2, e, A)$. Details to obtain maximum likelihood estimates via the EM algorithm are provided in Steibel et al. (2015). Our first posterior probability of interest is $\mathrm{pr}(G_{il} = g_{il}|X)$, which will be used for predicting the underlying genotype-ASE status of the lth SNP in the ith individual. Here X represents the RNA read counts for all n individuals at all L SNP positions. Given the states of G_{il} as defined in (4.1), we will also be able to deduce the ASE status for the respective individual and SNP. The posterior probability $\mathrm{pr}(G_{il}|X)$ can be computed by Bayes' formula. Let $G_i = (G_{i1}, \cdots, G_{iL})^{\mathrm{T}}$ represent all the possible genotype-ASE status combinations. Then the posterior probability is

$$L_{i,k}(l) := \mathrm{pr}(G_{il} = k|X) = \sum_{G_i} \mathrm{pr}(G_i|X) I(G_{il} = k) = \sum_{G_i} \frac{\mathrm{pr}(X, G_i)}{\mathrm{pr}(X)} I(G_{il} = k). \tag{4.6}$$

In order to estimate the hidden state of G_{il} ($i = 1, \ldots, n$; $l = 1, \ldots, L$) we compute $\max_k L_{i,k}(l)$ for each individual and SNP combination. The quantity $L_{i,k}(l)$ is computed from the EM algorithm. By the definition of the random variable δ_{il}, we aim to use the estimated state of G_{il} to obtain an estimate for the ratio of ASE. This leads to our second posterior probability of interest.

Let θ^* be the updated parameter vector upon convergence of the EM algorithm. Consider $f(\delta_{il}|X_{il}, G_{il} = k; \theta^*)$ such that

$$f(\delta_{il}|X_{il}, G_{il} = k; \theta^*) = \frac{f(X_{il}|\delta_{il}; \theta^*) f(\delta_{il}|G_{il} = k; \theta^*)}{\int f(X_{il}|\delta_{il}; \theta^*) f(\delta_{il}|G_{il} = k; \theta^*) d\delta_{il}}, \tag{4.7}$$

where the distributions of $f(X_{il}|\delta_{il}; \theta^*)$ and $f(\delta_{il}|G_{il} = k; \theta^*)$ are defined in (4.4) and (4.3), respectively. It follows that the denominator of (4.7) can be expressed as

$$f(X_{il}|G_{il} = k; \theta^*) = \begin{cases} \binom{n_l}{X_{il}}(1-e)^{X_{il1}}(\frac{e}{3})^{n_{il}-X_{il1}} & \text{for } k = 1, \\ \binom{n_l}{X_{il}}(0.5-\frac{e}{3})^{X_{il1}+X_{il4}}(\frac{e}{3})^{X_{il2}+X_{il3}} & \text{for } k = 2, \\ \binom{n_l}{X_{il}}(\frac{e}{3})^{X_{il2}+X_{il3}}\frac{C_0(\theta^*;X_{il1},X_{il4})}{0.5^{\alpha_1+\beta_1-1}B(\alpha_1,\beta_1)} & \text{for } k = 3, \\ \binom{n_l}{X_{il}}(\frac{e}{3})^{X_{il2}+X_{il3}}\frac{C_1(\theta^*;X_{il1},X_{il4})}{0.5^{\alpha_2+\beta_2-1}B(\alpha_2,\beta_2)} & \text{for } k = 4, \\ \binom{n_l}{X_{il}}(1-e)^{X_{il4}}(\frac{e}{3})^{n_{il}-X_{il4}} & \text{for } k = 5 \end{cases} \tag{4.8}$$

where $\binom{n_l}{X_{il}} = \frac{n_l!}{X_{il1}!X_{il2}!X_{il3}!X_{il4}!}$, and

$$C_0 = \int_{0.5}^{1} ((1-\frac{4e}{3})\delta_{il} + \frac{e}{3})^{X_{il1}}((\frac{4e}{3}-1)\delta_{il} + 1 - e)^{X_{il4}}(1-\delta_{il})^{\alpha_1-1}(\delta_{il}-0.5)^{\beta_1-1}d\delta_{il},$$

$$C_1 = \int_{0}^{0.5} ((1-\frac{4e}{3})\delta_{il} + \frac{e}{3})^{X_{il1}}((\frac{4e}{3}-1)\delta_{il} + 1 - e)^{X_{il4}}\delta_{il}^{\alpha_2-1}(0.5-\delta_{il})^{\beta_2-1}d\delta_{il}.$$

If $G_{il} = 3$ or $G_{il} = 4$, then we know that the heterozygous genotype has ASE with the quantity determined by a rescaled beta distribution. Thus, we only consider these estimated states when computing an estimate for δ_{il}. Hence, for $G_{il} = 3$ and $G_{il} = 4$ it follows that

$$f(\delta_{il}|X_{il}, G_{il} = k; \theta^*) = \begin{cases} M_0(\delta_{il}, X_{il1}, X_{il4}, \theta^*)(\delta_{il} - 0.5)^{\alpha_1 - 1}(1 - \delta_{il})^{\beta_1 - 1} & \text{for } k = 3, \\ \\ M_1(\delta_{il}, X_{il1}, X_{il4}, \theta^*)(\delta_{il})^{\alpha_2 - 1}(0.5 - \delta_{il})^{\beta_2 - 1} & \text{for } k = 4, \end{cases} \tag{4.9}$$

where

$$M_0 = ((1 - \frac{4e}{3})\delta_{il} + \frac{e}{3})^{X_{il1}}((\frac{4e}{3} - 1)\delta_{il} + 1 - e)^{X_{il4}} C_0(\boldsymbol{\theta^*}; X_{il1}, X_{il4})$$

$$M_1 = ((1 - \frac{4e}{3})\delta_{il} + \frac{e}{3})^{X_{il1}}((\frac{4e}{3} - 1)\delta_{il} + 1 - e)^{X_{il4}} C_1(\boldsymbol{\theta^*}; X_{il1}, X_{il4}),$$

with C_0 and C_1 as defined above.

We define our ASE ratio estimate as $\hat{\delta}_{il}$ ($i = 1, \ldots, n;\ l = 1, \ldots, L$), where $\hat{\delta}_{il}$ is the mode of the posterior distribution in (4.9).

4.3.2 Identification of quantitative trait loci

In order to quantify the impact of ASE on phenotypic variation we utilize $\hat{\delta}_{il}$ as estimated in Section 4.3.1. For L large, we aim to find a sparse solution for the L-dimensional parameter vector γ. In order to accomplish this task we apply a Lasso penalty. A sparse solution is computed via cyclical coordinate descent and k-fold cross validation. For Y_i as defined in (4.5) an estimated vector $\hat{\gamma}$ is given by

$$\hat{\gamma} = \arg\min_{\gamma} \left\{ -\sum_{i=1}^{n} \left[\frac{y_i \sum_{l=1}^{L} \hat{\delta}_{il}\gamma_l - b(\sum_{l=1}^{L} \hat{\delta}_{il}\gamma_l)}{a(\phi)} + c(y_i, \phi) \right] + \lambda^* ||\gamma||_1 \right\}, \tag{4.10}$$

where λ^* is a non-negative regularization parameter. If Y_i has a Binomial distribution, then a solution for $\hat{\gamma}$ is given by

$$\hat{\gamma} = \arg\min_{\hat{\gamma}} \left\{ -\sum_{i=1}^{n} (y_i \hat{\delta}_i^T \tilde{\gamma} - \log(1 + e^{\hat{\delta}_i^T \gamma})) + \lambda^* ||\gamma||_1 \right\}. \tag{4.11}$$

Similarly, if Y_i has as Gaussian distribution, then a solution for $\hat{\gamma}$ is given by

$$\hat{\gamma} = \arg\min_{\hat{\gamma}} \left\{ \sum_{i=1}^{n} (y_i - \hat{\delta}_i^{\,T} \gamma)^2 + \lambda^* ||\gamma||_1 \right\}. \tag{4.12}$$

A cyclic coordinate descent algorithm can be used to solve (4.10) – (4.12) (Friedman et al. 2010). Solutions are provided across a range of λ^* values, so to determine an optimal sparse solution we perform k-fold cross validation as a way to extract SNPs that have non-zero coefficients for a specific value of λ^*. Our choice of λ^* is based on the "one-standard error" rule as it provides the most parsimonious model whose error is no more than one standard error above the error of the best model. Details on cyclic coordinate descent and how it can be applied to specific exponential families is available in Friedman et al. (2010).

After identifying a sparse solution for $\tilde{\gamma}$ we apply ordinary least squares using the phenotypic response and filtered $\hat{\delta}$ to obtain estimates and standard errors for the non-zero γ_{S}. Ordinary least squares estimates are given by

$$\hat{\gamma}^* = \arg\min_{\gamma^*} \left\{ \sum_{i=1}^{n} \left(y_i - \hat{\delta}_i^{\,*\mathrm{T}} \gamma^* \right)^2 \right\}, \tag{4.13}$$

where $*$ denotes a filtered set of predictors and parameters after (4.10).

We summarize our procedure as follows. First, obtain ASE ratio estimates given RNA read counts and imputed genotype-ASE statuses. Second, apply variable selection to determine the SNPs with ASE that influence phenotypic variation. Third, model the relationship between SNPs with ASE and the response using ordinary least squares. Figure 4.2 illustrates the relationships between G_{il}, X_{il}, δ_{il}, and Y_i for the ith individual and $L = 5$.

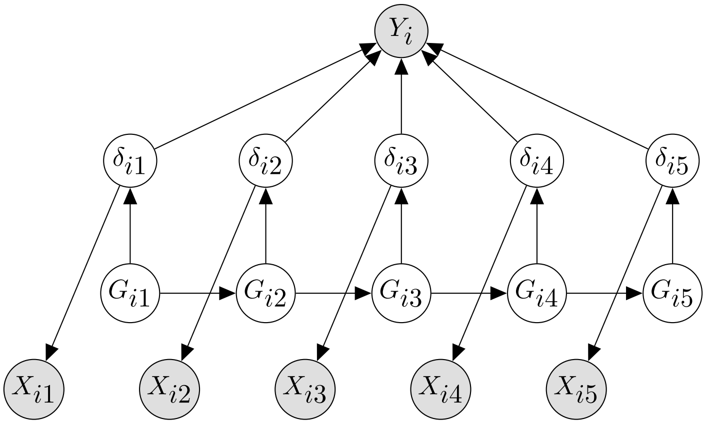

Figure 4.2: Grey circles represent observed values. White circles represent latent variables.

4.4 Simulation studies

In this section we demonstrate the performance of our proposed two-stage model in identifying significant SNPs with ASE ratios as they relate to a phenotype. We consider a simplified version of (4.1) for the simulation by ignoring an extended classification of heterozygous genotype-ASE states. Hence, we assume

$$G_{il} = \begin{cases} 1 & \text{for "AA"}, \\ 2 & \text{for "AT-ASE"}, \\ 3 & \text{for "TT"}, \end{cases} \tag{4.14}$$

follows a three-state Markov process. Our data generation process consisted of the following steps. First, two independent haplotypes were generated to form genotypes. The sequences for each of the haplotypes were created using linkage disequilibrium information. For each individual and SNP, total RNA read counts were generated from a negative binomial distribution with parameter λ and probability parameter $p = .40$. RNA read counts of A, C, G, and T were then generated according to the total number of RNA read counts and (4.14) –

(4.15). From (4.14) it follows that

$$\delta_{il}|G_{il} = k \sim \begin{cases} I_{\{\delta_{il}=1\}} & \text{for } k = 1, \\ \text{Beta}(\alpha, \beta) & \text{for } k = 2, \\ I_{\{\delta_{il}=0\}} & \text{for } k = 3. \end{cases} \tag{4.15}$$

Let δ be an $n \times L$ matrix that represents the true allele-specific expression ratios given the true underlying genotype-ASE statuses. For a given individual and SNP, where the underlying genotype is homozygous, the corresponding value in the matrix δ is represented by zero. We set these values to zero because our interest is only in exploring the cis-acting genetic effects on phenotypic variation.

In the next step we generate the phenotypic response by the following linear model, where

$$y_i = \sum_{l=1}^{L} \delta_{il}^{\mathrm{T}} \gamma_l + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{4.16}$$

and γ is an L-dimensional parameter vector. We assume γ is sparse and only allow the first four elements to be non-zero. Thus, $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, 0, \ldots, 0)^{\mathrm{T}}$ such that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4$. Under this set-up, the first four SNPs have cis-acting effects while the remaining $L - 4$ SNPs have no cis-acting effects on y_i.

In the simulation studies we set $n = 50$ and 100, and $L = 8, 15$, and 50. The parameter λ defined in the negative binomial distribution to simulate total RNA read counts was set to 16 and 24. The signal strength used in generation of the continuous phenotypic data,γ, was set to 2, 3, 5, and 7. Lastly, we set $e = 0.07$, $\alpha = 3$, $\beta = 3$ for the mapping parameter and Beta distribution parameters, respectively; and the linkage disequilibrium information was set to 0.30. The simulation results presented in the Tables are figures were based off 100 replications.

To evaluate the performance of our proposed method considered the false positive rate and false negative rate. For a given parameter combination, the false positive and false negative rates were averaged over the 100 replications. The false positive rate and false negative rate are defined as follows:

$$\begin{aligned} \text{False Positive Rate} &= \frac{\#\text{ coefficents falsely identified as non-zero}}{\#\text{ non-zero coefficients}} \\ \text{False Negative Rate} &= \frac{\#\text{ coefficents falsely identified as zero}}{\#\text{ zero coefficients}} \end{aligned} \tag{4.17}$$

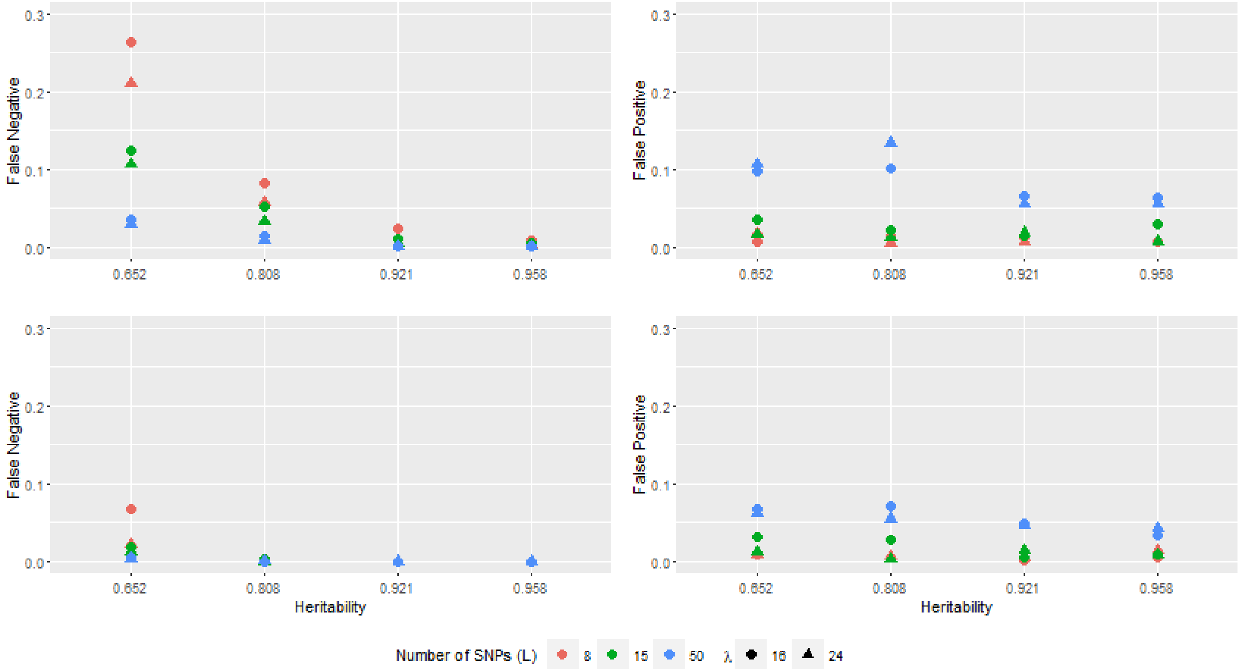

Figure 4.3: Average false negative rates and average false positive rates for the proposed method. Facets in row 1 are for $n = 50$. Facets in row 2 are for $n = 100$.

The simulation results for a single test at significance level 0.01 are illustrated in Figure 4.3. As the heritability, or value of γ increases, the average false negative rate decrease. The same relationship holds for the average false positive rates. As the number of SNPs increases for a given γ, the average false positive rate increases whereas the average false negative rate decreases. As the sample size increases both average rates decrease. The top row of plots corresponds to the setting in which $n = 50$, and the bottom row of plots corresponds to the setting in which $n = 100$. Lastly, all else held constant, a larger value of λ generally results in smaller false negative and false positive rates. A larger value of λ means a larger number of RNA read counts, and thus, more information. From a practical perspective, a low average false negative rate implies significant SNPs will not fail to be identified. Similar trends exist when we consider a simultaneous test. Average rate values are displayed in Table 4.1 and Table 4.2 for the single and simultaneous test, respectively.

Table 4.1: Average false positive and average false negative rates for the single test with significance level 0.01. Average false positive rate is top value

		n = 50				n = 100			
L	λ	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$
8	16	0.0074	0.0138	0.0120	0.0078	0.0085	0.0060	0.0020	0.0060
		0.2633	0.0828	0.0240	0.0085	0.0678	0.0040	0.0000	0.0000
8	24	0.0182	0.0040	0.0073	0.0060	0.0093	0.0060	0.0040	0.0133
		0.2105	0.0573	0.0040	0.0020	0.0220	0.0000	0.0000	0.0000
15	16	0.0357	0.0217	0.0145	0.0290	0.0308	0.0278	0.0060	0.0093
		0.1240	0.0515	0.0099	0.0052	0.0184	0.0034	0.0000	0.0000
15	24	0.0160	0.0120	0.0185	0.0073	0.0120	0.0020	0.0133	0.0080
		0.1062	0.0328	0.0042	0.0017	0.0117	0.0000	0.0000	0.0000
50	16	0.0977	0.1021	0.0650	0.0642	0.0676	0.0716	0.0486	0.0330
		0.0345	0.0138	0.0021	0.0021	0.0051	0.0002	0.0000	0.0000
50	24	0.1062	0.1337	0.0549	0.0562	0.0610	0.0545	0.0463	0.0420
		0.0288	0.0086	0.0011	0.0004	0.0034	0.0002	0.0000	0.0000

Table 4.2: Average false positive and average false negative rates for the simultaneous test with nominal level 0.05. Average false positive rate is top value

		n = 50				n = 100			
L	λ	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$
8	16	0.0718	0.0726	0.0702	0.0603	0.0512	0.0433	0.0293	0.0343
		0.0668	0.0213	0.0040	0.0020	0.0120	0.0000	0.0000	0.0000
8	24	0.0756	0.0854	0.0762	0.0450	0.0548	0.0400	0.0326	0.0363
		0.0440	0.0040	0.0000	0.0000	0.0065	0.0000	0.0000	0.0000
15	16	0.1233	0.1132	0.1177	0.1146	0.0975	0.0709	0.0378	0.0360
		0.0436	0.0106	0.0008	0.0009	0.0025	0.0000	0.0000	0.0000
15	24	0.1348	0.1141	0.0861	0.0885	0.0543	0.0273	0.0352	0.0327
		0.0350	0.0077	0.0027	0.0000	0.0025	0.0000	0.0000	0.0000
50	16	0.2033	0.2445	0.2238	0.2001	0.1544	0.1528	0.0925	0.0905
		0.0178	0.0043	0.0011	0.0013	0.0017	0.0002	0.0000	0.0000
50	24	0.2433	0.2884	0.1923	0.1848	0.1403	0.1264	0.1080	0.0884
		0.0124	0.0020	0.0004	0.0002	0.0015	0.0000	0.0000	0.0000

We evaluated the performance of our proposed method with two alternative procedures.

Alternative procedure 1 used an exact binomial test based on the simulated RNA read counts in order to estimate the unknown genotype-ASE status. Our test was performed under the null hypothesis $p = 0.50$ with alternatives $p > 0.50$ and $p < 0.50$ corresponding to genotypes AA and TT, respectively. Following genotype-ASE state imputation, we performed an ordinary least squares post Lasso technique using the simulated phenotypic data as the response and the estimated genotype-ASE statuses as predictors. Thus, we did not consider ASE estimation in this alternative procedure. The average false positive and average false negative rates were calculated under the same parameter scenarios as our proposed method.

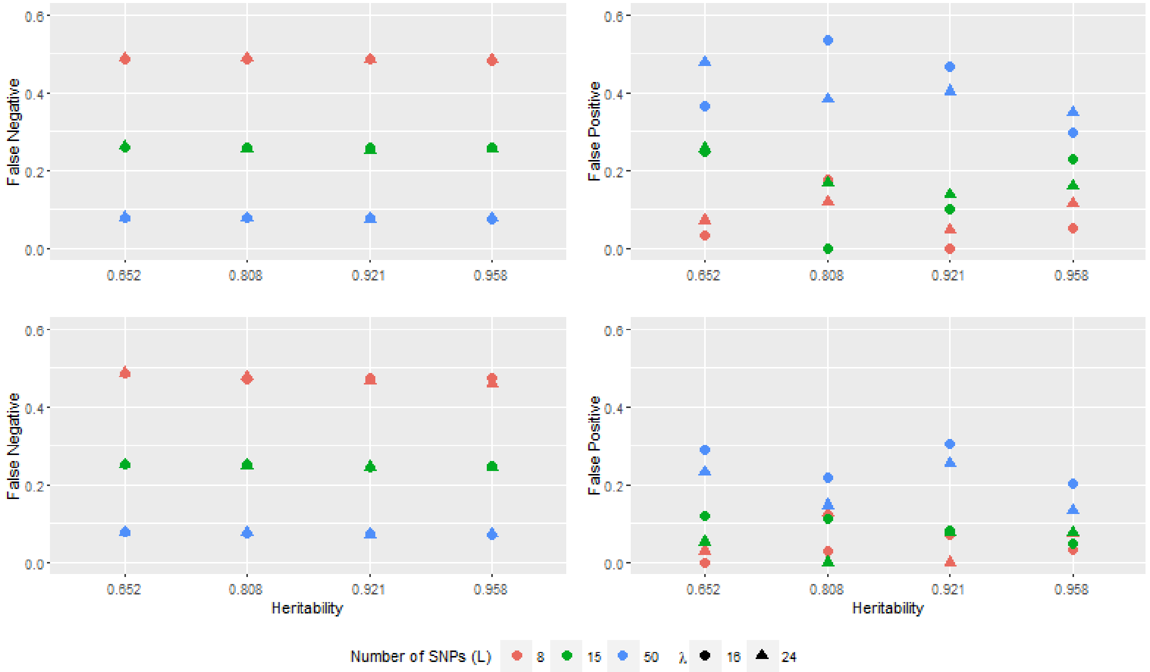

Figure 4.4: Average false negative rates and average false positive rates for alternative procedure 1. Facets in row 1 are for $n = 50$. Facets in row 2 are for $n = 100$.

Figure 4.4 depicts the average false positive and average false negative rates. When n increases from 50 to 100, the average false negative rate decreases slightly. However, we do not see the precipitous decline in average false negative rates as heritability increases, compared to our proposed method. Likewise, the average false positive rate decreases as the sample size increases. As the number of SNPs increases, the average false positive rate increases and is much higher than the average rates in Figure 4.3. Table 4.3 provides the raw values for all parameter combinations.

Table 4.3: Alternative method 1, average false positive and average false negative rates for the single test with significance level 0.01. Average false positive rate is top value

		n = 50				n = 100			
L	λ	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$
8	16	0.0312	0.1750	0.0000	0.0500	0.0000	0.0309	0.0714	0.0333
		0.4864	0.4869	0.4867	0.4843	0.4864	0.4726	0.4752	0.4746
8	24	0.0714	0.1190	0.0476	0.1136	0.0294	0.1250	0.0000	0.0750
		0.4886	0.4884	0.4845	0.4850	0.4876	0.4776	0.4674	0.4578
15	16	0.2500	0.0000	0.1000	0.2304	0.1190	0.1133	0.0808	0.0469
		0.2592	0.2577	0.2587	0.2579	0.2528	0.2513	0.2434	0.2469
15	24	0.2564	0.1667	0.1369	0.1591	0.0526	0.0000	0.0778	0.0783
		0.2608	0.2555	0.2519	0.2538	0.2509	0.2493	0.2447	0.2424
50	16	0.3646	0.5370	0.4674	0.2978	0.2892	0.2179	0.3053	0.2048
		0.0778	0.0782	0.0763	0.0753	0.0767	0.0740	0.0757	0.0719
50	24	0.4769	0.3833	0.4031	0.3476	0.2325	0.1471	0.2546	0.1333
		0.0782	0.0775	0.0750	0.0760	0.0762	0.0757	0.0714	0.0723

The weak performance of alternative method 1 comes from two sources. First, the binomial test results in less accurate genotype predictions compared to (4.6). Assuming genotypes follow a Markov process and using a hidden Markov model to impute their state provides extra information that results in accurate predictions (Ferguson-Smith 2001). Second, correctly predicted heterozygous states do not consider ASE.

For alternative method 2 we estimated an allele-specific expression ratio directly from the simulated RNA read counts. Let $A\hat{S}E_{il}$ be an estimated quantity for ASE such that

$$\mathrm{A\hat{S}E}_{il} = \frac{X_{il_{ref}}}{X_{il_{ref}} + X_{il_{alt}}} \qquad i = 1, \ldots, n;\ l = 1, \ldots, L, \tag{4.18}$$

where we define A to be the reference allele and T to be the alternative allele. Again, we performed an ordinary least squares post-Lasso procedure in conjunction. The average false positive and average false negative rates were calculated under the same parameter scenarios as our proposed method and alternative method 1. Figure 4.5 and Table 4.4 illustrate that the performance is similar to alternative method 1 and inferior to our proposed method with regards to false positive and false negative metrics.

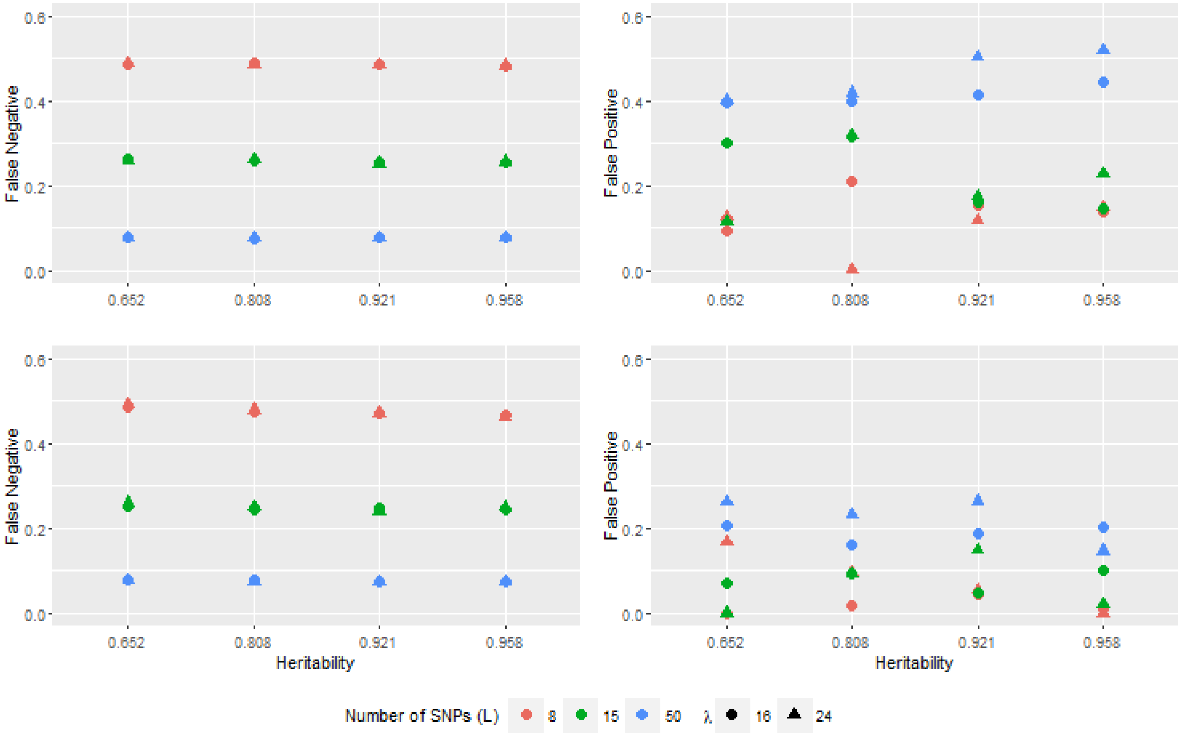

Figure 4.5: Average false negative rates and average false positive rates for alternative procedure 2. Facets in row 1 are for $n = 50$. Facets in row 2 are for $n = 100$.

Table 4.4: Alternative method 2, average false positive and average false negative rates for the single test with significance level 0.01. Average false positive rate is top value

		n = 50				n = 100			
L	λ	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 7$
8	16	0.0938	0.2111	0.1522	0.1391	0.0000	0.0185	0.0455	0.0095
		0.4875	0.4917	0.4867	0.4817	0.4860	0.4738	0.4710	0.4680
8	24	0.1250	0.0000	0.1190	0.1481	0.1667	0.0962	0.0556	0.0000
		0.4876	0.4850	0.4861	0.4832	0.4924	0.4793	0.4716	0.4623
15	16	0.3000	0.3148	0.1618	0.1458	0.0702	0.0938	0.0500	0.1000
		0.2628	0.2614	0.2552	0.2552	0.2529	0.2443	0.2490	0.2455
15	24	0.1154	0.3182	0.1746	0.2283	0.0000	0.0909	0.1496	0.0208
		0.2591	0.2626	0.2521	0.2562	0.2598	0.2517	0.2399	0.2499
50	16	0.3968	0.4011	0.4141	0.4429	0.2083	0.1607	0.1865	0.2033
		0.0787	0.0756	0.0770	0.0778	0.0773	0.0773	0.0746	0.0745
50	24	0.4009	0.4190	0.5042	0.5198	0.2633	0.2333	0.2646	0.1470
		0.0771	0.0777	0.0773	0.0758	0.0765	0.0737	0.0732	0.0720

Figure 4.6 characterizes the discrepancy between an ASE estimate from RNA read counts defined in (4.18) and an ASE estimate using our proposed hierarchical model. For values less than 0.50, the hidden Markov model ASE estimate is greater than our naive estimate in (4.18). Above 0.50 the hidden Markov model ASE estimate is less than our naive estimate.

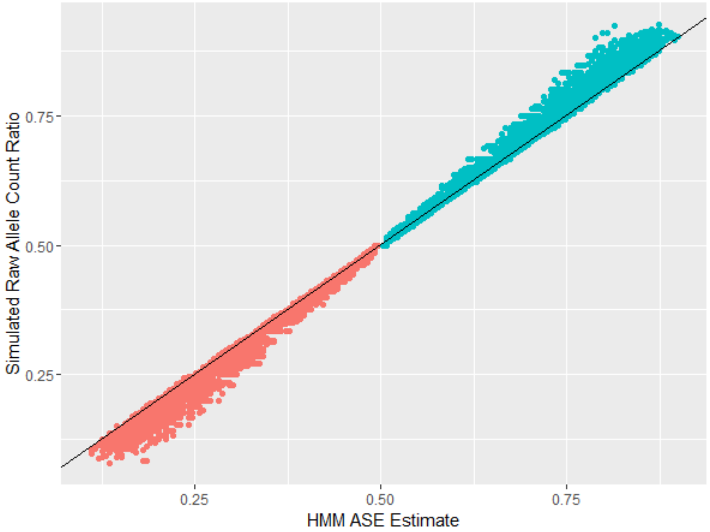

Figure 4.6: ASE estimates from the hidden Markov model compared to simulated raw allele count ratios. Hidden Markov model imputed ASE ratios with value less than 0.50 are marked as red, and values above 0.50 are marked as blue.

Through our simulation analysis, our hierarchical model appears to perform better than two alternative procedures at identifying SNPs with cis-acting effects on phenotypic variation.

4.5 An empirical study

The following paragraph provides some of the data gathering and processing details as explained in Steibel et al. (2015). RNA sequence data was obtained from 24 female pigs from an F2 cross of Duroc and Pietrain breeds (Choi et al. 2012, Choi et al. 2011, Edwards et al. 2008a, Edwards et al. 2008b, Steibel et al. 2011). Protocols for RNA sequencing and the accuracy of genotype calling using a hidden Markov-ASE model have already been established in Steibel et al. (2015). To summarize the process, RNA from each sample was reverse transcribed, fragmented, barcode-labeled and sequenced on an Illumina HiSeq 2000 (100 bp, paired-end reads). After quality control filtering, sequence reads were aligned to reference genome (Sus scrofa 10.2.69 retrieved from the Ensembl database) using Tophat (Trapnell et al. 2009). Coding SNP discovery and genotyping were done with VarScan (Trapnell et al. 2009). We focused on chromosome 13 and extracted counts of reads agreeing with reference (R) or alternative (A) allele with respect to the reference genome at putative 5364 cSNP and we retained read counts on 65 SNPs that could be independently validated using a SNP chip (Steibel et al. 2015). In addition to the RNA sequence data, 45 minute post-mortem meat pH was recorded in these animals (Edwards et al. 2008b) and served as our phenotypic response variable for analyses.

The RNA sequence data we analyzed is available in the HMMASE R package which is available at `http://www.stt.\msu.edu/users/pszhong/HMMASE.html`. The data set was partitioned so that the minimum number of SNPs in a segment is 30. Our proposed procedure was applied to each segment of RNA sequence data. Figure 4.7 depicts estimates for significant SNPs along with their SNP ID number. For example, the second segmented data set produced four significant SNPs: 12256008, 12400307, 12403644, and 12404379.

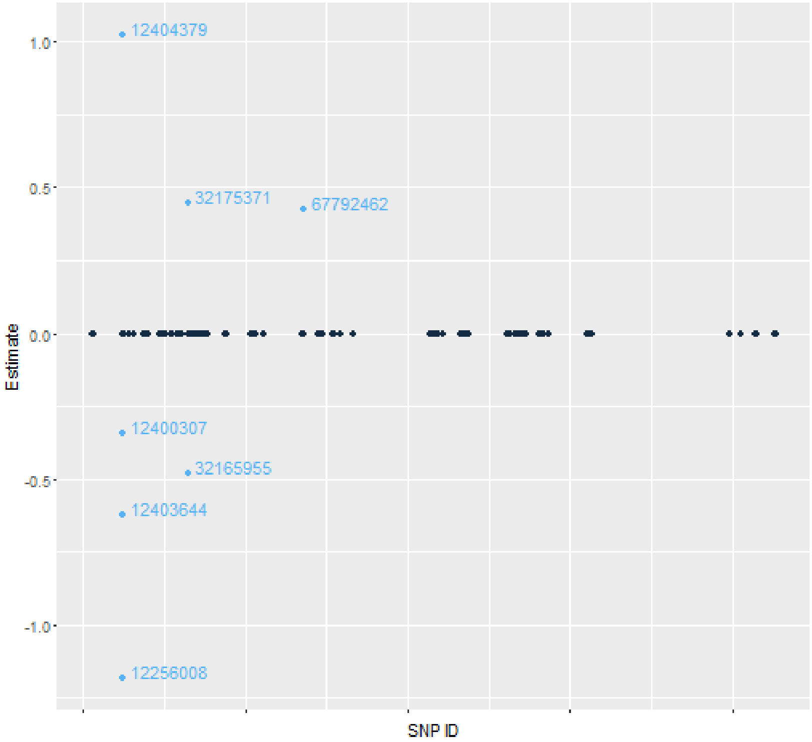

Figure 4.7: Estimates for SNPs. Significant SNPs are displayed with their respective ID provided in the real data set. IDs correspond to the ordered locations.

We investigated the effects of an estimated ASE ratio from the hidden Markov model compared to the naive estimate defined in (4.18). Figure 4.8 depicts the relationship between the two estimates, and reveals a shrinkage around each Beta distribution's mode of 0.25 and 0.75, respectively. For values below the respective mode, the hidden Markov ASE estimate is less than the raw allele count ratio, and for values above the respective mode, the hidden Markov ASE estimate is greater than the raw allele count ratio.

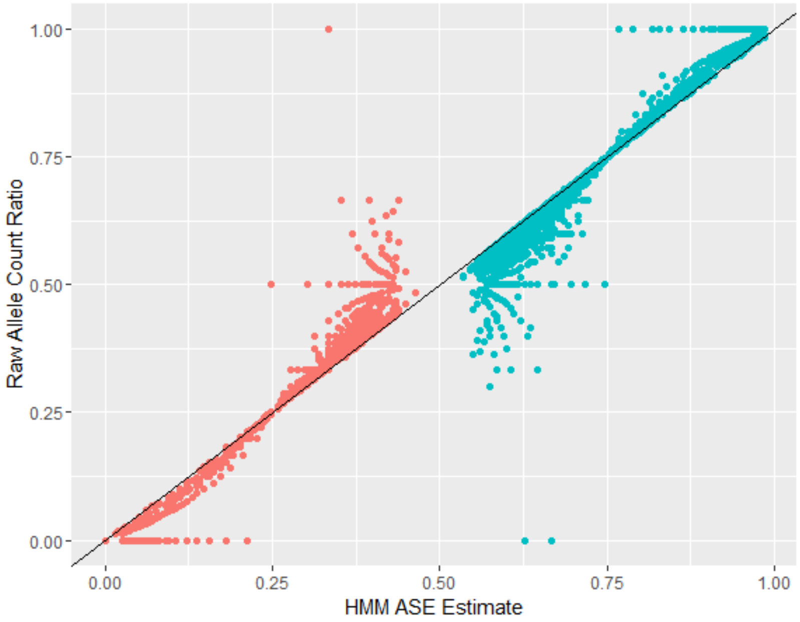

Figure 4.8: ASE estimates from the hidden Markov model compared to real raw allele count ratios. Hidden Markov imputed ASE values conditional on $G_{il} = 3$ and $G_{il} = 4$ are marked as blue and red, respectively.

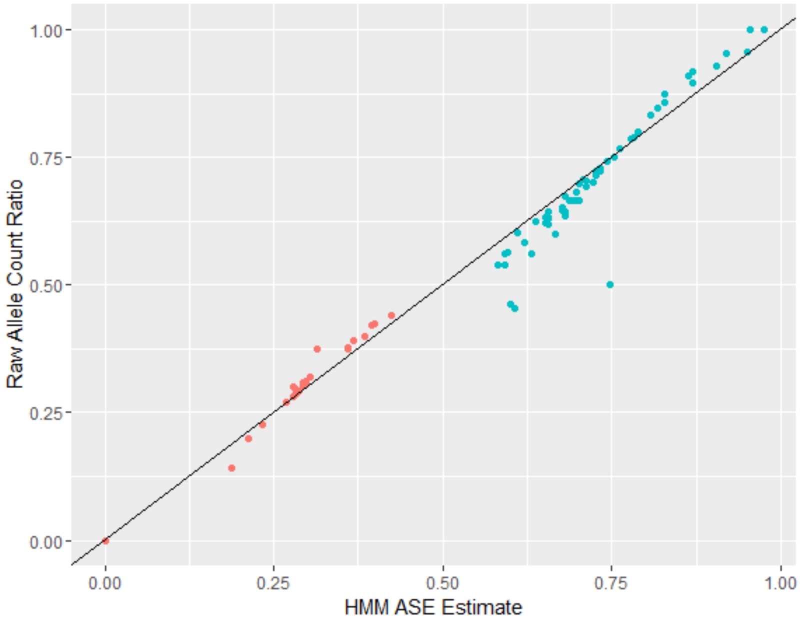

Figure 4.9: ASE estimates from the hidden Markov model compared to real raw allele count ratios. Hidden Markov imputed ASE values conditional on $G_{il} = 3$ and $G_{il} = 4$ are marked as blue and red, respectively.

CHAPTER 5

CONCLUSION

5.1 Introduction

In this chapter we summarize the salient contributions to the field of Statistics based on the content of Chapters 2 through 4. We also introduce details to new and exciting research challenges.

5.2 Summary of contributions

In Chapter 2, we proposed a novel nonparametric test procedure for testing the temporal homogeneity of covariance matrices with high-dimensional longitudinal data. The procedure aims to detect and identify change points among a temporally dependent collection of covariance matrices. In Chapter 2, a new test statistic was introduced, and theoretical results were derived under an asymptotic setting in which n and p diverge and T is finite. The test statistic's asymptotic distribution was derived under mild dependence assumptions but with no assumption of sparsity and no requirement on the relationship between n and p. We also proposed a procedure to identify the locations of change points through binary segmentation. The corresponding change point identification estimator's rate of convergence was investigated and shown to be consistent provided an adequate signal-to-noise ratio exists. Numerical studies demonstrated the finite sample performance of our procedure. These developments expanded the field of Statistics by pioneering a robust procedure to detect and identify change points among covariance matrices in the presence of high-dimensional longitudinal data.

In Chapter 3, we widened the scope of applicability with regards to the procedure developed in Chapter 2. Theoretical results were derived under an asymptotic framework in which n, p, and T all diverge. We established the test statistic's asymptotic distribution and demonstrated that the change point identification estimator's rate of convergence depends on n, p, T, and the signal-to-noise ratio. Therefore, the estimator was also shown to be consistent in a diverging T setting, provided an adequate signal-to-noise ratio exists. Numerical studies demonstrated the finite sample performance for a large T setting. Chapter 3 also addressed computation challenges. Recursive formulae were derived as were computation efficient forms of U-type statistics. In addition, we proposed an accurate quantile approximation procedure to via an estimated correlation matrix. The overall computation complexity was reduced from the order pn^4T^6 to the order of pn^2T^3. These theoretical and computational developments of Chapter 3 made our procedure applicable to high-dimensional functional data and allowed us to demonstrate our method using a task-based fMRI data set. Thus, the contribution to Statistics in Chapter 3 is an expanded scope of the cutting-edge procedures introduced in Chapter 2.

In Chapter 4, we developed a hierarchical model to understand the relationship between allele-specific expression and phenotypic variation. Our hierarchical model was able to use RNA sequence data and identify SNPs with ASE that have a cis-acting effect on a phenotypic response. The performance is accurate and can quickly be applied through a combination of the EM algorithm and Lasso procedure.

5.3 Future research

The procedures established in Chapter 2 and extended in Chapter 3 required mild assumptions but did not allow much flexibility in terms of n and T. For example, in many longitudinal studies patients drop out, measurements are missing at random or non-random time points, and the sample size can be extremely small or even one. The methods developed in Chapters 2 and 3 will not be applicable to data under these settings. More work is necessary to accommodate a wider domain of real-world data and problems. One valuable extension of our work will be to develop a procedure for single-subject inference in high-dimensional longitudinal data and high-dimensional functional data. In terms of a homogeneity test for covariance matrices, this setting will allow for an increase in scope of applications due to more realistic assumptions, and a sample size requirement of only one. For genetic or fMRI data, an effectively developed procedure will invoke a personalized medicine approach and have a greater benefit to the individual patient. Other potential applications of this work would include real estate and financial data, and motion sensor data for activities.

From a computation standpoint, accurate and fast approximations could be developed to handle situations where T is of the order 1000. Even with a high-performance computing cluster, it is not practical to apply our proposed procedure for massive values of T. However, as technology improves and longitudinal studies expand, the demand to address massive high-dimensional longitudinal data will increase. It is paramount that statistical methods can produce accurate and fast results for practitioners.

A natural extension to the model proposed in Chapter 4 is to develop a unified likelihood approach in a hierarchical framework. Rather than only use RNA read count data to predict the underlying genotype with ASE status, we could perform this prediction given phenotypic data and RNA read counts. The additional information should improve prediction accuracy. From a theoretical perspective, a unified likelihood approach would allow for statistical inference, and under certain conditions consistency and asymptotic normality could be proved. From a computation perspective, we could investigate a procedure that performs variable selection and parameter estimation through a penalized EM algorithm or penalized variational EM algorithm.

BIBLIOGRAPHY

BIBLIOGRAPHY

AHMAD, R. M. (2017). Testing homogeneity of several covariance matrices and multi-sample sphericity for high-dimensional data under non-normality. *Communications in Statistics - Theory and Methods*, 46:8, 3738–3753.

AMINIKHANGHAHI, S. & COOK, D. (2016). A survey of methods for time series change point detection. *Knowledge And Information Systems*, 51(2), 339–367.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley.

ASHBURNER, M., BALL, C., BLAKE, J., BOTSTEIN, D., BUTLER, H., & CHERRY, J. ET AL. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nature Genetics,* 25, 25–29.

AUE, A., HORMANN, S., HORVÁTH, L., & REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics,* 37, 4046–4087.

BACH, F. R. (2008). Bolasso: Model consistent lasso estimation through the bootstrap. In Cohen, W. W., Mccallum, A., & Roweis, S. T., editors, *Proceedings of the 25th International Conference on Machine Learning*, pages 33–40, Brookline, MA. Microtome Publishing.

BAI, Z. & SARANADASA, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica,* 6, 311–329.

BAI, Z. D. & YIN, Y. Q. (1993). Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix. *Annals of Probability*, 21, 1276–1294.

BALDASSANO, C., CHEN, J., ZADBOOD, A., PILLOW, J., HASSON, U., & NORMAN, K. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, 95(3), 709–721.e5.

BARNETT, I. & ONNELA, J.-P. (2016). Change point detection in correlation networks. *Scientific Reports,* 6, 18893.

BASSEVILLE M. & NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes - theory and application.* Prentice Hall.

BOX, G. E. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika,* 36, 317–346.

BRODSKY, B. E. (2017). *Change-Point Analysis in Nonstationary Stochastic Models,* CRC Press.

BRODSKY, B. E. & DARKHOVSKY, B.S. (1993). *Nonparametric Methods in Change Point Problems*, Boston: Springer.

Brown, R., Durbin, J. & Evans, J. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society: Series B*, 37(2), 149–192.

Buckland, P. (2004). Allele-specific gene expression differences in humans. *Human Molecular Genetics,* 13 (suppl 2), R255–R260.

Bühlmann, P. & van de Geer, S. (2011). *Statistics for High-Dimensional Data*, Boston: Springer.

Cai, T., Liu, W., & Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501), 265–277.

Cai, T. & Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B), 2359–2388.

Cai, T. & Xia, Y. (2014). High-dimensional sparse MANOVA. *Journal of Multivariate Analysis*, 131, 174–196.

Candes, E. & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35(6), 2313–2351.

Chen, J., Leong, Y., Honey, C., Yong, C., Norman, K., & Hasson, U. (2016). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20(1), 115–125.

Chen, S. X. & Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Annals of Statistics,* 38, 808–835.

Chen, S. X., Zhang, L., & Zhong, P.-S. (2010). Testing high dimensional covariance matrices. *Journal of the American Statistical Association,* 105, 810–819.

Cheng, H., Perumbakkam, S., Pyrkosz, A., Dunn, J., Legarra, A., & Muir, W. (2015). Fine mapping of QTL and genomic prediction using allele-specific expression SNPs demonstrates that the complex trait of genetic resistance to Marek's disease is predominantly determined by transcriptional regulation. *BMC Genomics*, 16(1).

Chernoff, H. & Zacks, S. (1964). Estimating the Current Mean of a Normal Distribution which is Subjected to Changes in Time. *Annals of Mathematical Statistics*, 35(3), 999–1018.

Choi, I., Bates, R., Raney, N., Steibel, J., & Ernst, C. (2012). Evaluation of QTL for carcass merit and meat quality traits in a US commercial Duroc population. *Meat Science*, 92:132–8.

Choi, I., Steibel, J., Bates, R., Raney, N., Rumph, J., & Ernst, C. (2011). Identification of Carcass and Meat Quality QTL in an F(2) Duroc x Pietrain pig resource population using different least-squares analysis models. *Frontiers in Genetics*, 2:18.

Csörgö, M. & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*, New York: John Wiley.

Danaher, P., Paul, D., & Wang, P. (2015). Covariance-based analyses of biological pathways. *Biometrika,* 102, 533–544.

de la Chapelle, A. (2009). Genetic predisposition to human disease: allele-specific expression and low-penetrance regulatory loci. *Oncogene*, 28(38), 3345–3348.

Dempster, A. (1958). A High Dimensional Two Sample Significance Test. *Annals of Mathematical Statistics*, 29(4), 995–1010.

Dette, H., Pan, G., & Yang, Q. (2018). Estimating a change point in a sequence of very high-dimensional covariance matrices. Arxiv.org.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *In AMS Conference on Mathematical Challenges of the 21st Century.*

Donoho, D. L. & Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3), 425–455.

Edwards, D., Ernst, C., Raney, N., Doumit, M., Hoge, M., & Bates, R. (2008a). Quantitative trait loci mapping in an F2 Duroc x Pietrain resource population: I. Growth traits. *Journal of Animal Science*, 86:241–53.

Edwards, D., Ernst, C., Raney, N., Doumit, M., Hoge, M., & Bates, R. (2008b). Quantitative trait locus mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. *Journal of Animal Science*, 86:254–66.

Efron, B. (2007). Size, power and false discovery rates. *Annals of Statistics,* 35, 1351–1377.

Fan, J., Han, F., & Liu, H. (2014a). Challenges of Big Data analysis. *National Science Review*, 1(2):293–314.

Fan, J. & Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96:456, 1348–1360

Fan, J. & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. In Sanz-Sole, M., Soria, J., Varona, J. L., & Verdera, J., editors, *Proceedings of the International Congress of Mathematicians*, volume 3, pages 595–622, Zürich, CH. European Mathematical Society.

Fan, J., Liao, Y., & Yao, J. (2015). Power enhancement in high dimensional cross-sectional tests. *Econometrica,* 83, 1497–1541.

Ferguson-Smith, A. (2001). Imprinting and the Epigenetic Asymmetry Between Parental Genomes. *Science,* 293(5532), 1086–1089.

Finn, E., Shen, X., Scheinost, D., Rosenberg, M., Huang, J., & Chun, M. et al. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11), 1664–1671.

FRIEDMAN, J., HASTIE, T., & TIBSHIRANI, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software,* 33(1), 1–22.

FU, W. & KNIGHT, K. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 28(5):1356–1378.

FUJIKOSHI, Y., ULYANOV, V., & SHIMIZU, R. (2010). *Multivariate Statistics*, Wiley Series In Probability And Statistics.

GENZ, A., BRETZ, F., MIWA, T., MI, X., LEISCH, F., SCHEIPL, F., & HOTHORN, T. (2018). *mvtnorm: Multivariate Normal and t Distributions,* R package version 1.0–8.

GU, F. & WANG, X. (2015). Analysis of allele specific expression - a survey. *Tsinghua Science And Technology*, 20(5), 513–529.

HALL, P. & HEYDE, C. (1980). *Martingale Limit Theory and Its Application*, New York: Academic Press.

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika,* 57, 1–17.

HORVÁTH, L. & KOKOSZKA, P. (1997). The effect of long-range dependence on change-point estimators. *Journal of Statistical Planning And Inference*, 64(1), 57–81.

HUBER, P. J. (1981). *Robust statistics*, New York: John Wiley.

JOHNSON, R. & BAGSHAW, M. (1974). The Effect of Serial Correlation on the Performance of CUSUM Tests. *Technometrics*, 16(1), 103–112.

JOHNSTONE, I. (2001). On the distribution of the largest eigenvalue in principal components analysis.*Annals of Statistics,* 29 (2001), no. 2, 295–327.

JOHNSTONE, I. & TITTERINGTON, D. (2009). Statistical challenges of high-dimensional data. *Philosophical Transactions of The Royal Society A: Mathematical, Physical And Engineering Sciences*, 367(1906), 4237–4253.

KANDER, Z. & ZACKS, S. (1966). Test Procedures for Possible Changes in Parameters of Statistical Distributions Occurring at Unknown Time Points. *Annals of Mathematical Statistics*, 37(5), 1196–1210.

KANNAN, R. P., HENSLEY, L. L., EVERS, L. E., LEMON, S. M., & MCGIVERN, D. R. (2011). Hepatitis C virus infection causes cell cycle arrest at the level of initiation of mitosis. *Journal of Virology,* 85, 7989–8001.

KOH, W., PAN, W., GAWAD, C., FAN, H. C., KERCHNER, G. A., WYSS-CORAY, T., BLUMENFELD, Y. J., EL-SAYED, Y. Y., & QUAKE, S. R. (2014). Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proceedings of the National Academy of Sciences,* 111, 7361–7366.

Kundu, S., Ming, J., Pierce, J., McDowell, J., & Guo, Y. (2018). Estimating dynamic brain functional networks using multi-subject fMRI data. *Neuroimage*, 183, 635–649.

Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Berg, J. J., Greene, D. J., McCarthy, J. E., Tagliazucchi, E., Laufs, H., Schlaggar, B. L., Dosenbach, N. U. F., & Peterson, S. E. (2017). On the stability of BOLD fMRI correlations. *Cerebral Cortex,* 27, 4719–4732 .

Ledoit, O. & Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Annals of Statistics*, 30(4):1081–1102.

Li, J. & Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Annals of Statistics,* 40, 908–940.

Meinshausen, N. & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B*, 72(4):417–473.

Monti, R., Hellyer, P., Sharp, D., Leech, R., Anagnostopoulos, C., & Montana, G. (2014). Estimating time-varying brain connectivity networks from functional MRI time series. *Neuroimage*, 103, 427–443.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*, New York: John Wiley.

Nariai, N., Kojima, K., Mimori, T., Kawai, Y., & Nagasaki, M. (2016). A Bayesian approach for estimating allele-specific expression from RNA-Seq data with diploid genomes. *BMC Genomics*, 17(S1).

Osborne, M. R., Presnell, B., & Turlach, B. A. (2000). On the LASSO and its dual. *Journal of Computational and Graphical Statistics*, 9(2):319–337.

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, 41(1–2), 100–115.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika,* 47, 379–396.

Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis,* New York: Springer.

Schapiro, A., Rogers, T., Cordova, N., Turk-Browne, N., & Botvinick, M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, 16(4), 486–492.

Schott, J. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample size. *Computational Statistics & Data Analysis,* 51, 6535–6542.

Sen, A. & Srivastava, M. (1973). On Multivariate Tests for Detecting Change in Mean. *Sankhyā: The Indian Journal of Statistics: Series A*, 35(2), 173–186.

Shedden, K. & Taylor, J. (2005). Differential correlation detects complex associations between gene expression and clinical outcomes in lung adenocarcinomas. *Methods of Microarray Data Analysis*, J. S. Shoemaker and S. M. Lin, eds. Boston: Springer, pp 121–131.

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*, 82, 403–415.

Skelly, D., Johansson, M., Madeoy, J., Wakefield, J., & Akey, J. (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Research*, 21(10), 1728–1737.

Srivastava, M. S. & Worsley, K. J. (1986). Likelihood Ratio Tests for a Change in the Multivariate Normal Mean. *Journal of the American Statistical Association*, 81:393, 199–204.

Srivastava, M. S. & Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis,* 101, 1319–1329.

Steibel, J., Bates, R., Rosa, G., Tempelman, R., Rilington, V., & Ragavendran, A. et al. (2011). Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. *PLOS One*, 6(2), e16766.

Steibel, J., Wang, H., & Zhong, P.-S. (2015). A hidden Markov approach for ascertaining cSNP genotypes from RNA sequence data in the presence of allelic imbalance by exploiting linkage disequilibrium. *BMC Bioinformatics*, 16(1).

Storey, J., Xiao, W., Leek, J., Tompkins, R., & Davis, R. (2005). Significance analysis of time course microarray experiments. *Proceedings of the National Academy of Sciences,* 102, 12837–12842.

Tai, Y. & Speed, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *Annals of Statistics,* 34, 2387–2412.

Taylor, M., Tsukahara, T., Brodsky, L., Schaley, J., Sanda, C., & Stephens, M. et al. (2007). Changes in gene expression during pegylated interferon and ribavirin therapy of chronic hepatitis C virus distinguish responders from nonresponders to antiviral therapy. *Journal of Virology,* 81, 3391–3401.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B*, 67(1):91–108.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9):1105–11.

Venkatraman, E. S. (1992). Consistency results in multiple change-point situations. *Technical report*, Department of Statistics. Stanford University.

Wang, D., Yu, Y., & Rinaldo, A. (2017). Optimal Covariance Change Point Localization in High Dimension. Arxiv.org.

Yang, Q. & Pan, G. (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *Journal of the American Statistical Association,* 112, 188–200.

Yao, Y. & Davis, R. (1986). The Asymptotic Behavior of the Likelihood Ratio Statistic for Testing a Shift in Mean in a Sequence of Independent Normal Variates. *Sankhyā: The Indian Journal of Statistics: Series A*, 48(3), 339–353.

Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1):49–67.

Zacks, J., Speer, N., Swallow, K., Braver, T., & Reynolds, J. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133(2), 273–293.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., & Breakspear, M. (2014). Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences,* 111, 10341–10346.

Zhang, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38(2), 894–942.

Zhang, C., Bai, Z., Hu, J., & Wang, C. (2018). Multi-sample test for high-dimensional covariance matrices. *Communications in Statistics - Theory and Methods*, 47:13, 3161–3177.

Zhang, J. & Boos, D. D. (1992). Bootstrap critical values for testing homogeneity of covariance matrices. *Journal of the American Statistical Association,* 87, 425–429.

Zhang, C. & Zhang, S. (2013). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of The Royal Statistical Society: Series B*, 76(1), 217–242.

Zhao, P. & Yu, B. (2006). On Model Selection Consistency of Lasso. *Journal of Machine Learning Research*, 2541–2563.

Zheng, S., Bai, Z., & Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Annals of Statistics,* 43, 546–591.

Zhu, L.-X., Ng, K., & Jing, P. (1992). Resampling methods for homogeneity tests of covariance matrices. *Statistica Sinica,* 12, 769–783.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320.